INSTITUTO POLITÉCNICO DE BEJA

Escola Superior de Tecnologia e Gestão

Mestrado em Engenharia de Segurança Informática

iKNOW – Sistema distribuído de *intelligence* em Fontes Abertas

Hélder Filipe Dias Antão Tomás

Beja

2019

INSTITUTO POLITÉCNICO DE BEJA

Escola Superior de Tecnologia e Gestão

Mestrado em Engenharia de Segurança Informática

iKNOW – Sistema distribuído de *intelligence* em Fontes Abertas

Elaborado por:
Hélder Filipe Dias Antão Tomás

Orientado por:
Professor Doutor Rui Miguel Soares Silva

Dissertação de Mestrado, apresentada na Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Beja

2019

Resumo

As informações (no sentido de *intelligence*) têm desempenhado desde sempre, um papel crítico nas organizações governamentais, privadas, empresariais, militares, … Com as informações certas no tempo certo, ganham-se guerras, garante-se a Defesa, melhora-se a qualidade de vida dos cidadãos e ajuda-se as empresas a preparar-se contra a sua concorrência (*competitive intelligence*), entre tantas outras utilizações.

Os Governos de cada país e as estruturas militares têm conhecimento disto e têm apostado milhões na procura e tratamento de informações/*intelligence*, nas suas mais variadas formas. Hoje, a vivermos na Idade do Conhecimento, temos a Internet e milhões de pessoas a divulgar a todo o momento, informações sobre si e sobre o seu meio envolvente, quer em redes sociais, quer em blogues, quer nos mais diversos meios de comunicação "abertos". Acessíveis a qualquer pessoa.

Pretendeu-se nesta dissertação, apresentar o estado de arte de um dos ramos da *intelligence*, a *Open Source Intelligence* (OSINT), que permite obter informações sem infringir a Lei, aprofundar as suas vantagens e desvantagens comparativamente a outras fontes de obtenção de informação (incluindo as fechadas). Nomeadamente, na disponibilidade e no preço.

Pretendeu-se paralelamente a esta dissertação, desenvolver um sistema informático distribuído, assente em diversas tecnologias e metodologias, a operar sobre a Internet, para recolha automatizada de dados em fontes abertas, tratamento da informação e obtenção de inteligência, gerando também alarmística, e informação processada designada normalmente por *threat intel*, útil em segurança informática e física, quer de uma organização, empresa, pessoa, quer de um Estado.

O sistema criado, iKNOW, opera em equipamentos baratos, recolhe informação em sites, fóruns, redes sociais e motores de busca, consoante palavras-chave/cartões de crédito/links… actuando em várias profundidades de URL's, assim como na rede TOR. Gera relatórios cifrados com chave pública e envia-os a quem criou a operação. O site criado para o efeito também oferece outras ferramentas e métricas. O sistema consiste em um ou mais servidores (*nGinx*, *PHP*, *MySQL*, *Bash*, *Python*) e muitos clientes (preferência caiu nos *Raspberrys* com *Python*, *Bash*, e *scripts* próprios). Cada cliente pode criar a sua operação, mas estas são distribuídas e executadas por qualquer um, com tempo definido, *user-agent* alterado, aleatoriedade na execução, entre outros. Relatórios são enviados por email, cifrados com chave assimétrica RSA. São gerados *analytics* do obtido e enviados como *threat intelligence* para consumidores desta informação (SIEM, SOC, CSIRT, emails…).

Os clientes (máquinas/sensores) são monitorizados em operacionalidade, disponibilidade, hora local, e temperatura em tempo "quase" real.

Palavras-chave: informações, *osint*, *open source intelligence*, *intelligence*, segurança, espionagem, *Deep Web*, anonimização, iKNOW, sistema distribuído, cibercrime

Abstract

iKNOW – Intelligence system in Open Sources

Intelligence has always played a critical role in government, private, business, security, defense and military organizations. With the right information in the right time, wars have been won, Defense was guaranteed, and the quality life of general citizens, improved. Intelligence also helps companies prepare themselves against their competitors (competitive intelligence,) among many other uses.

The governments of each country and the military structures are aware of this and have invested millions in seeking and processing information (to generate intelligence) in its most varied forms. Today, living in the "Age of Knowledge", we have the Internet and millions of people spreading information about themselves and their surroundings at all times, whether on social networks, blogs, or the most diverse "open "media. All this information accessible to anyone.

The aim of this dissertation is to present the state of the art of one of the branches of intelligence, Open Source Intelligence or OSINT, which allows obtaining information without breaking the law or resorting to espionage, to deepen its advantages and disadvantages compared to other sources of information gathering. Namely, availability and price.

In parallel to this dissertation, it was intended to develop a distributed computer system, to show how OSINT gathering and goals can be achieved with anonymity and security. Several technologies were used, operating on the Internet, with the goal of (semi) automated data gathering in open sources, information processing and intelligence generation, generating alarmistic and *threat intel,* useful in both computer and physical security, as well as in organizations, people and state/government.

The system created, iKNOW, operates on cheap equipment, collects information on websites, forums, (social networks *not anymore*) and major search engines, depending on keywords/credit cards/links… acting at various URL's depths, as well as in the TOR/deep web environment. Generates public key encrypted reports and sends them to the person who created the operation. The purpose-built site also offers other tools and metrics. The hardware is based on a server (nGinx, PHP, MySQL, Bash, Python) and many clients (preference dropped on Raspberrys with Python, Bash, and own scripts). Each client can create their operation, but they are distributed and executed by anyone, with a defined time, altered user-agent, randomness of execution, among others. Reports are sent by email, encrypted with RSA asymmetric key. The machines / sensors are monitored for (almost) real time operation, availability, local time, and temperature. Analytics are generated from the obtained and sent as threat intelligence to consumers of this information (SIEM, SOC, CSIRT, emails, …).

Keywords: Intelligence, osint, open source intelligence, security, Deep Web, espionage, anonymization, iKNOW, information, open source, distributed system, cibercrime

Agradecimentos

A realização desta dissertação de mestrado demorou alguns anos e contou com apoios e incentivos de várias pessoas. Muito aconteceu, muito mudou.
Ao meu orientador, Professor Doutor Rui Silva, agradeço o apoio e orientação no trabalho, para transformar “um livro” numa dissertação, e pela oportunidade dada de poder ter escrito sobre este tema.
À Susana, pelo tempo e abnegação por tantas coisas durante estes anos.
Ao Rogério, pelo companheirismo, apoio e confiança.
Aos pais e familiares, pela preocupação e incentivo constantes para que a obra iniciada pudesse um dia ter fim.

Acrónimos e Abreviaturas

BYOD	*Bring Your Own Device*
CIA	*Central Intelligence Agency*
CISMIL	Centro de Informações e Segurança Militar
SOC	Centro de Operações de Segurança (tipicamente informática)
CNCS	Centro Nacional de CiberSegurança
CERT	*Computer Emergency Response Team*
DoS	*Denial of Service* - Nome atribuído a um ataque informático de negação de serviço, que impede o funcionamento ou que outros utilizadores tenham acesso a esse serviço.
DDoS	*Distributed Denial of Service* - DoS, como referido acima, mas executado por diversos utilizadores. Tipicamente muitos.
DINFO	Divisão de Informação
DIMIL	Divisão de Informações Militares
EMGFA	Estado-Maior-General das Forças Armadas
ENISA	*European Network and Information Security Agency*
GEOINT	*Geospatial Intelligence*
HUMINT	*Human Intelligence*
IC	Infraestrutura Crítica
NSA	*National Security Agency.*
NATO	*North Atlantic Treaty Organization*
OSINT	*Open Source Intelligence*
PJ	Polícia Judiciária, polícia de investigação criminal portuguesa
RGPD	Regulamento Geral de Privacidade e Protecção de Dados Pessoais
SSH	S*ecure Shell*, Protocolo de comunicação cifrada em rede.
SIS	Serviço de Informações de Segurança
SIED	Serviço de Informações Estratégicas de Defesa
SVR	Serviços de informações externas russas. Sucedeu ao KGB e trata de informações não-militares.
SIGINT	*Signals Intelligence*
SIRP	Sistema de Informações da República Portuguesa
SSI	Sistema de Segurança Interna

Índice Geral

Índice de Figuras

CAPÍTULO I - Introdução

1.1 Motivação e Enquadramento

As informações têm desempenhado desde sempre, um papel crítico nas organizações governamentais, privadas, empresariais… Com as informações correctas ganham-se guerras, garante-se a Defesa, melhora-se a qualidade de vida dos cidadãos e ajuda-se as empresas a preparar-se contra a sua concorrência (*competitive intelligence*), entre tantas outras utilizações.

Os Governos de cada país e as estruturas militares têm conhecimento disto e têm apostado milhões na procura e tratamento de informações, nas suas mais variadas formas. Hoje, a vivermos na Idade do Conhecimento, temos a Internet e milhões de pessoas a divulgar a todo o momento, informações sobre si e o seu meio envolvente.

OSINT (*Open source intelligence*) é o acrónimo usado, para descrever a inteligência[1], no sentido de informações, como em serviços de inteligência/informações. É obtida através de dados disponíveis para o público em geral, como jornais, artigos académicos, revistas científicas emissões de TV, mapas e documentos online, mesmo que em fóruns e sites que exijam um *login* de utilizador. Não é espionagem e não pode utilizar métodos de força bruta ou que de alguma forma, contornem defesas técnicas (sejam elas físicas ou lógicas (informáticas/electrónicas, …)).

O conhecimento aqui estudado, *Open Source Intelligence* – OSINT vai incidir sobre as informações/*intelligence* em fontes abertas. O que são, como podem ser utilizadas, vantagens estratégicas e competitivas, sendo que as maiores residem na disponibilidade (estão disponíveis em todo o lado), e no custo (geralmente de forma gratuita) sendo que também não faz uso de técnicas (padrão) de espionagem nem (no seu geral) vai contra a Lei. São abordados os fenómenos da *Deep Web*, anonimização na rede assim como sua utilização e riscos associados, monitorização da rede com vista à vigilância da actividade dos mais novos, compra e venda de forma anónima, investigação, correlação de eventos, entre outros.

Pretendeu-se também com esta Dissertação, o desenvolvimento de um sistema que utilize as potencialidades do OSINT para automatizar recolhas (automáticas) de informação de diversas fontes conforme determinado como "alvo". O objectivo é recolher informação (texto, imagens, vídeo, coordenadas GPS, localização, palavras-chave, entre outros) e compilar depois a informação num relatório que deve ser enviado imediatamente a quem o pediu, logo que algum dos pontos-chave seja detectado, gerando no fim um relatório final. O relatório produzido

[1]O termo Inteligência, foi adoptado da palavra inglesa *intelligence* do contexto militar, e que depois foi adaptado a outras ciências e meios, como por exemplo pelas polícias e pela economia e negócios. O termo inteligência ou *intelligence* apoia as organizações na tomada de decisões que podem ser estratégicas sobre a competição/concorrência ou inimigos. O termo pode ser abreviado por *intel*. Hoje em dia, já se ouvem também outros termos como *business intelligence*. Fonte: http://whatis.techtarget.com/definition/ELINT-electronic-intelligence

tem uma estrutura ponderada pelas diversas recomendações de quem pediu a pesquisa e deu o alvo. Este sistema utiliza no seu funcionamento as linguagens de programação *Python*, *Bash*, *PHP*, bases de dados *MySQL*/*MariaDB*, entre outros.

Os contributos que poderão ser trazidos por esta ferramenta e estudo, estendem-se às áreas das informações, forças de segurança, órgãos de investigação criminal, cidadão "comum" mas acima de tudo, aos analistas que pretendem numa base diária, obter informações de forma automática, medir o pulso da comunidade relativamente a um dado problema, antever problemas/perigos e evitá-los. O código pode ser melhorado e adaptado, tendo sido comentado e criado de forma a simplificar a sua utilização e a ser compreendido.

O trabalho que se apresenta está dividido em seis capítulos sendo que se pretende apresentar no primeiro, a introdução e apresentação sumária do que se pretendeu.

No segundo capítulo, conceitos básicos de informações/intelligence, o actual estado da arte, a Lei Portuguesa para crimes como espionagem, e a forma como o OSINT existe e é aplicado pelas nossas forças militares, investigação e segurança, assim como pelos *hackers*. Pretendeu-se também mostrar as vantagens e desvantagens, assim como formas de nos escondermos e defendermos.

O terceiro capítulo, fala sobre o que nos propomos fazer. A hipótese de investigação, o iKNOW, suas potencialidades, e algumas técnicas que se propõe utilizar.

O quarto capítulo, pretende mostrar como foi implementado o sistema iKNOW, seus requisitos, como pode ser reconstruído, funcionamento, arquitectura, evolução do 0.1 ao actual e falhas.

O quinto capítulo, avaliação, pretende dar a conhecer o que foi tido como objectivos, o que se conseguiu e o que se pode melhorar. Contém o resultado de inquéritos e avaliação por terceiros. Contém ainda a metodologia OSINT e a forma como o iKNOW a utilizou.

Por fim, o capitulo seis, conclusões, contém um sumário de todo o trabalho efectuado, uma pequena comparação entre fontes abertas e fontes fechadas e humanas(espionagem). Conta com um pequeno sub-capítulo que mostra o resultado final do iKNOW em números e imagens.

De forma a ser conseguida uma estrutura de trabalho de acordo com os *standards* académicos para este tipo e nível de estudos, foram colocados nos apêndices muitas informações e tópicos de relevo no âmbito OSINT e iKNOW que não poderiam ser colocados em volume de informação. Deste modo, serão remetidos para os "Apêndices" sempre que necessário ou de interesse.

CAPÍTULO II - Estado da Arte

2.1 Estado da Arte

Segundo o Finslab[2], o termo OSINT é mundialmente reconhecido e associado a *Open Source Intelligence*, e significa inteligência (ou informações) obtida através do recurso a fontes disponíveis publicamente. Na comunidade de "*Intelligence*" o termo "aberto" ou *open*, refere-se a fontes de informação disponíveis publicamente. Não está relacionada com "inteligência pública" ou a *software* de código aberto.

Segundo Pedro Borges Graça[3], OSINT é um conceito "*que se encontra em expansão no âmbito dos serviços de informações e que globalmente traduz a aplicação às fontes abertas da metodologia empregue na produção das informações confidenciais*".

Ainda segundo o Jornal de Negócios, o conceito de OSINT não existia praticamente há 15 anos. Não constituía uma opção real para se obter informações (boas), dado que, ainda segundo o mesmo site "um dos manuais de referência da época, «*Perfectly Legal Competitor Intelligence*», de *Douglas Bernhardt*, editado pelo *Finantial Times* em 1993, a afirmação é que «o campo das fontes abertas é infinito, os analistas não têm capacidade nem tempo para as tratar e, por isso, do ponto de vista da produção de informações o seu valor qualitativo é baixo»".

Desde então, várias entidades de renome mudaram a sua atitude relativamente ao OSINT. Por exemplo, a CIA[4], anunciou a criação de um departamento especializado em OSINT. Também a NATO em 2001 definiu o conceito de "*open source data*" ((OSD) fotos e imagens de satélite comerciais) e "*open source information*" ((OSI), meios de comunicação social, livros, relatórios), sendo ambas referentes a informações em bruto (sem tratamento).

Segundo a NATO, a OSINT é «a informação que foi deliberadamente descoberta, discriminada, destilada e disseminada por uma audiência seleccionada, de modo a responder a uma questão específica».

A definição é consensual (a distinção do que é e de como pode ser utilizada a OSINT nem por isso...) e foi colocada em prática em diversos domínios, tais como na guerra da informação, guerra de inteligência competitiva e económica.

Relativamente a Portugal, dizia o Jornal Económico em 2005, que Portugal não tinha esta sensibilidade ao nível das empresas enquanto factor de vantagem competitiva. Que não existem unidades de informações estratégicas e que a procura de vantagens em informação está a cargo de amadores da "navegação". Se houver dinheiro, compram-se simplesmente as informações, sem filtro nem estudo adequado das mesmas.

"…*A Internet é um elemento central neste contexto e em Portugal parece não existir ainda sensibilidade ao nível das empresas para o seu potencial* enquanto factor de vantagem

[2]http://finslab.com/enciclopedia/letra-a/abrir-inteligencia-fonte.php, 28-04-2015

[3]http://www.jornaldenegocios.pt/opiniao/detalhe/o_novo_conceito_de_osint.html, 18 Novembro 2005, 13:59 por Pedro Borges Graça

[4]Central Intelligence Agency – CIA, é a agência de espionagem dos Estados Unidos. Uma das mais reputadas a nível mundial

competitiva via OSINT." As unidades de informações estratégicas são inexistentes *e geralmente a procura de dados e informações na Internet, no seio das empresas, fica a cargo dos habilidosos e amadores da «navegação» ou de alguém designado para o efeito, acumulando melhor ou pior, casuisticamente, informação bruta em folhas de papel em dossiês. Quando há dinheiro, compra-se informações externamente sob a forma de relatórios, que são muitas vezes peças caras de «pronto-a-decidir», sem qualquer filtro consistente de conhecimento interno, estratégico, cientificamente produzido…[5]".*

Este mesmo *site* referenciava também a necessidade de as pesquisas terem obrigatoriamente de ter um «regimento», orientação e técnica, dando os exemplos dos motores de busca especializados e cada vez mais eficazes. Pelo lado negativo, especificava a acção dos *hackers* no OSINT, como a potencial desvantagem competitiva.

A prática de OSINT não pode ser feita de forma amadora ou poderemos perder-nos num mar de informação e a maior parte do tempo, "lixo". Tem de ter orientação e técnica, e esta mesma indicação é-nos dada por diversos promotores de OSINT (Robert Steele por exemplo) e a metodologia OSINT, se seguida, fornece esta mesma orientação.

Em 2017, uma das notícias mais faladas no ano, foram as eleições americanas, tendo ganho Donald Trump quando nada o fazia prever. Muito se falou, (especialmente quando o mesmo começou a falar) de *fake news*, o que no mundo das informações se entende como propaganda e/ou contra-informação. Veio, entretanto, a público que na verdade, houve mesmo "*fake news*" e que estas vieram a ajudar Trump nas eleições (e não o contrário). As notícias falsas, explicadas pelo jornal New York Times[6] espalhadas nas redes sociais, vistas e "consumidas" por supostamente milhões de pessoas, poderão ter influenciado os votantes na hora da escolha. A suposta origem? Rússia. Objectivo? Não tanto Trump ganhar, mas sim Hillary Clinton perder. Desde *leaks* de emails a *posts* falsos, valeu tudo. "*The site's phony promoters were in the vanguard of a cyberarmy of counterfeit Facebook and Twitter accounts, a legion of Russian-controlled impostors whose operations are still being unraveled*." Também muito clarificador, *"…Facebook and Twitter, the American companies that essentially invented the tools of social media and, in this case, did not stop them from being turned into engines of deception and propaganda…"*

A luta e a utilização da propaganda foram uma das razões pelas quais a CIA foi criada.

> "*A ditadura perfeita terá a aparência da democracia, uma prisão sem muros na qual os prisioneiros não sonharão sequer com a fuga. Um sistema de escravatura onde, graças ao consumo e ao divertimento, os escravos terão amor à sua escravidão.*"
>
> Aldous Huxley.

Aldous Huxley disse também "*Sessenta e duas mil repetições fazem uma verdade*.". Um dizer popular para "uma mentira dita mil vezes torna-se verdade".

[5]http://www.jornaldenegocios.pt/opiniao/detalhe/o-novo-conceito-de-osint

[6]New York Times, *The Fake Americans Russia Created to Influence the Election*" online em https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html

"On Twitter, as on Facebook, Russian fingerprints are on hundreds or thousands of fake accounts that regularly posted anti-Clinton messages."

Scott Shane, New York Times[7]

2.2 Conceitos básicos de informações/*intelligence*

A informação pode ser definida como o conhecimento no seu estado mais bruto. Para ser útil (portanto valiosa), precisa ser transformada, colocada num ambiente, numa data e num local, ser explicada e ser simples porque uma transmissão deficiente não ajuda a tomar uma decisão, e é nisso que a *intelligence* se destaca.

De forma muito resumida (pois será novamente retomado o tema), existem dois conceitos importantes a ponto de justificar este capítulo e de nos referirmos a informações como *intelligence*. De salientar que até recentemente era utilizado nas informações portuguesas, o termo "informações estratégicas, não se utilizando o termo de inteligência ou *intelligence*, isto devido a serem termos que derivaram do âmbito militar, de onde nasceram.
A Inteligência/*intelligence pode ser definida como informação*:

- que é passível de ser compreendida;
- com valor acrescentado;
- que foi analisada no contexto da sua força e considerada confiável;

A análise da informação pode ser definida como:

- a resolução ou separação de uma informação nas partes que a compõem;
- a determinação dessas partes;
- o rastreamento dessas partes até à sua origem de forma a descobrir os princípios gerais que lhe deram origem, assim como a sua localização e tempo;
- a conclusão dos resultados deste processo;

Daí que,

Informação + sua avaliação/análise = *Intelligence*

De uma forma simples, a análise de inteligência resume-se ao recolher de informação, avaliá-la sumariamente, seguindo-se um processamento e análise, para assim obter dados úteis/inteligência. Essas conclusões serão depois utilizadas para ajudar ao auxílio de decisões ponderadas.

A *intelligence* é de tal forma importante em determinados países que são das primeiras

[7] The Fake Americans Russia Created to Influence the Election, disponível online em https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html

etapas do dia-a-dia desses Governos. Por exemplo, nos Estados Unidos da América, o Presidente é informado logo de manhã, pelo *Director of National Intelligence*[8] no que respeita às relações internacionais. Na Rússia, o presidente Putin é aconselhado, entre outros, com o Director-Geral do SVR (antigo KGB). Note-se então a relevância das informações que tanto o Governo dos EUA como da Rússia, estão cercados ao nível do topo, de elementos dos serviços de informações do respectivo país.

2.3 Geração e classificação de informações pela Defesa e forças militares

"Um exército sem agentes secretos é um homem cego e surdo"
Sun Tzu, in "A Arte da Guerra"

A Defesa e as forças militares são dos maiores interessados em informações/*intelligence*. Foi na defesa que nasceram muitos dos termos relativos a informações[9], usados hoje em dia não só nas Forças Armadas mas também pelas forças de segurança e sociedade civil. A Defesa é obrigada a fazer o melhor uso possível das informações/intelligence de forma a ser o mais eficaz possível, já que um erro pode colocar vidas humanas em risco e a perda da soberania nacional. Serão de seguida introduzidas várias noções de informações/intelligence e a forma como a Defesa trata estes assuntos, muito semelhantes ao da sociedade civil mas com uma importância maior.

Noção, Distinção e Classificação

Os militares consideraram que as informações não são todas iguais e como tal, depois de distinguidas, classificaram-nas. Notícia por exemplo, é segundo o "Manual de Informações"[10], "qualquer facto, documento ou material susceptível de contribuir para um melhor conhecimento do inimigo actual ou potencial, ou da área de operações", com uma classificação que varia em função do seu grau de confiança (confiança na fonte, possibilidade de ser verdadeira comparando o que se conhece da notícia com o conhecimento da própria pessoa). Pode depois ser processada e arquivada para uso posterior, ou não.
O termo "informação", é utilizado pelos militares como sendo o resultado/conhecimento obtido em função de um estudo (pesquisa, estudo e interpretação) de todas as notícias relacionadas com o assunto.
Já para "informações", a noção envolve o conhecimento de informação, mas de uma forma extrapolativa, que envolve saber porque aconteceu, por quem, se pode acontecer de novo e quando. Estas "informações" também conhecidas por *intelligence* são usadas neste trabalho como forma de distinção de informações no âmbito não-militar.

[8]Estudos de Intelligence, Pedro Borges Graça, pág.22
[9] Ver apêndice "1. OSINT – Evolução histórica"
[10] Manual de Informações, EME, página 6

Classificação das informações quanto à utilização

As informações criadas pela Defesa, podem classificar-se em estratégicas, operacionais ou tácticas, consoante a sua utilização posterior. São definidas como:

- Estratégicas, aquelas que dão origem a decisões estratégicas políticas e militares. O seu destino são o Governo e Ministérios.
- Operacionais, aquelas que são recolhidas e necessárias para o planeamento de operações. Tem como destino, os comandos militares.
- Tácticas, aquelas que são recolhidas e projectadas para o desenrolar imediato de operações. Têm como destino os líderes/comandantes no terreno.

Qualquer uma das informações acima, tem em conta a máxima premissa de que só deve ter acesso à informação quem realmente possa precisar dela para cumprir a sua missão. Esta premissa é também válida para a segurança da informação em geral.

Consoante a ameaça ("*qualquer acontecimento ou acção em curso ou previsível que contraria a consecução de um objectivo e que normalmente é causadora de danos materiais ou morais*[11]"), a Defesa, que é composta por diversas forças militares deve adoptar a melhor resposta para ser o mais eficaz possível. Por exemplo, uma missão que envolva mar vai com certeza incluir a Marinha. Uma missão que inclua a exfiltração de tropas ameaças ou cercadas vai envolver a força aérea, possivelmente Comandos ou Rangers de Operações Especiais, quiçá também Marinha. Diferentes problemas requerem diferentes abordagens e estratégias, daí também a opção das informações OSINT ser tão importante.

Uma missão de paz, que traga tropas de um país amigo vai com certeza ser melhor recebida do que tropas de um país remoto ou com quem tenha havido disputas no passado.

A Defesa/Estado-Maior General das Forças Armadas, tem também no seu interior, diversos organismos que procedem à colecta e disseminação de informações, ainda que apenas interiormente e no seu âmbito, como por exemplo, o Centro de Informações e Segurança Militares (CISMIL). À unidade que trata das informações já foram dados diversos nomes tais como: 2ª Divisão do EMGFA/DINFO/SIED/SIM/DIMIL. Actualmente a produção de informações militares está a cargo do SIEDM (produção de informações ao nível estratégico) e da DIMIL.

Nos apêndices (ponto 4) é possível ver esta informação, e alguns gráficos relativos à constituição e organização da Defesa Portuguesa.

2.4 Ramificações da intelligence - O OSINT

"*Se conheces o inimigo e te conheces a ti mesmo, não precisas temer o resultado de cem batalhas. Se te conheces, mas não conheces o inimigo, para cada vitória ganha sofrerás também uma derrota. Se não te conheces nem ao teu inimigo, perderás todas as batalhas...*"

[11] Noção de ameaça: Gen Abel Cabral Couto, Elementos de Estratégia, I Vol, Edição do IAEM, 1988, página 328.

Os termos e expressões da *intelligence*, derivam[12] [13] quase todos de instituições militares, onde já eram usados, e são depois adoptados pelas instituições civis, forças de segurança, etc .

É de comum acordo a existência de diversos ramos da *intelligence*. O acrónimo OSINT é um deles. Pode num futuro próximo, serem adicionados novos ramos devido à evolução tos tempos e tecnologias, sociedade, etc.

Resumidamente fica uma descrição de cada, sendo que tal pode ser considerado importante na escolha entre diversos tipos de intelligence, sendo que quase todas são de cariz fechado (não OSINT).

- HUMINT (*Human Intelligence*) – recolha de informações por fontes humanas, como serviços de informações (agentes/espiões), polícia (agentes infiltrados), entre outros (*insiders*/sabotadores/…). Segundo o FBI[14] americano, é dos mais importantes, a par das SIGINT e OSINT.
- IMINT (*Imagery Intelligence ou photo intel*ligence) – Recolha de informações através da utilização da fotografia e imagem como forma de conhecimento sobre o meio que rodeia. Foi muito utilizada durante a Guerra Civil americana, em balões tripulados por soldados, e na 2ª Guerra Mundial através de fotos aéreas para conhecimento das posições inimigas e criação de mapas.
- MASINT (*Measurement and Signatures Intelligence*) – Recolha de informações sobre actividades industriais e capacidades bélicas. É utilizada em conjunto com TELINT (*intelligence* de telemetria, sub-ramo da ELINT), ELINT, IMINT e SIGINT para obter dados à distância sobre armamento, potencial de destruição, tamanho, peso, emissões e seu tipo, entre outras informações de interesse. É um ramo de inteligência pouco falado embora seja hoje bastante utilizado (mas discretamente) pela Defesa. Lembrar o recente caso dos mísseis da Coreia do Norte e há pouco tempo, aquando das potenciais armas bioquímicas no Iraque.
- ELINT (*Electronic Intelligence*) – A ELINT teve o seu início na Segunda Guerra Mundial com a invenção do radar. Segundo a definição[15] da NSA[16], trata-se da recolha de informações por via de sinais electrónicos que não contêm voz/discurso ou texto (que para estes efeitos já se trataria de COMINT). Divide-se em TechELINT e OpELINT.
    - TechELINT, que se ocupa das características e funções dos sistemas de armas e sistemas de emissão de sinais como radares, bloqueadores de sinais e comunicações e sinais de navegação com o objectivo de criação de detecção de radares, contramedidas e equipamentos "anti-armas". Guerra electrónica,
    - OpELINT, por sua vez ocupa-se da localização de alvos da ELINT, auxiliando estes produtos os comandantes no campo de batalha.

[12] RBRAVO em "O Conceito de "Fontes Abertas " na Investigação do Cibercrime", "A maior parte das metodologias policiais (técnicas) têm raiz do meio militar… Com o tempo, a terminologia "sigint", "humint", "comint", "elint" e "osint" foi também incorporada no léxico policial, principalmente a partir de meados dos anos setenta nos USA e Canadá, e a partir dos anos oitenta na Europa, quando se intensificou de forma sistemática a capacitação das polícias criminais com a formação de analistas criminais…" PP2

[13] Vide apêndice "1.OSINT – Evolução histórica"

[14]https://www.fbi.gov/about-us/intelligence/disciplines

[15]https://www.nsa.gov/about/_files/cryptologic_heritage/publications/misc/elint.pdf

[16]Agência de Segurança Nacional norte-americana. Disponível online em https://www.nsa.gov

- SIGINT (*Signals Intelligence*) – Intercepção de transmissões electrónicas recolhidas através de aviões, satélites, embarcações marítimas, entre outros...
    - COMINT (*Communications intelligence*) - captura de informações de voz, texto e transmissões de sinais. Tem como objectivo determinar: quem participa nas comunicações, onde, quanto tempo, tipo de cifra utilizada, etc.
    - FISINT (*Foreign Instrumentation Signals INTelligence)*, tem como objectivo a intercepção das diversas formas e emissões electromagnéticas de entidades externas ao país, como aeronaves, embarcações, submarinos, entre outros.

Fica uma imagem recolhida do documento *Millitary intelligence*, apresentação do Tenente General Vizela Cardoso, e que representa de forma bastante interessante, a obtenção, o processamento e a distribuição das informações tratadas e agora úteis e englobadas num contexto e objectivo preciso, a quem delas precisa. Neste caso para militares, e com todos os ramos de *intelligence* a dar o seu contributo.

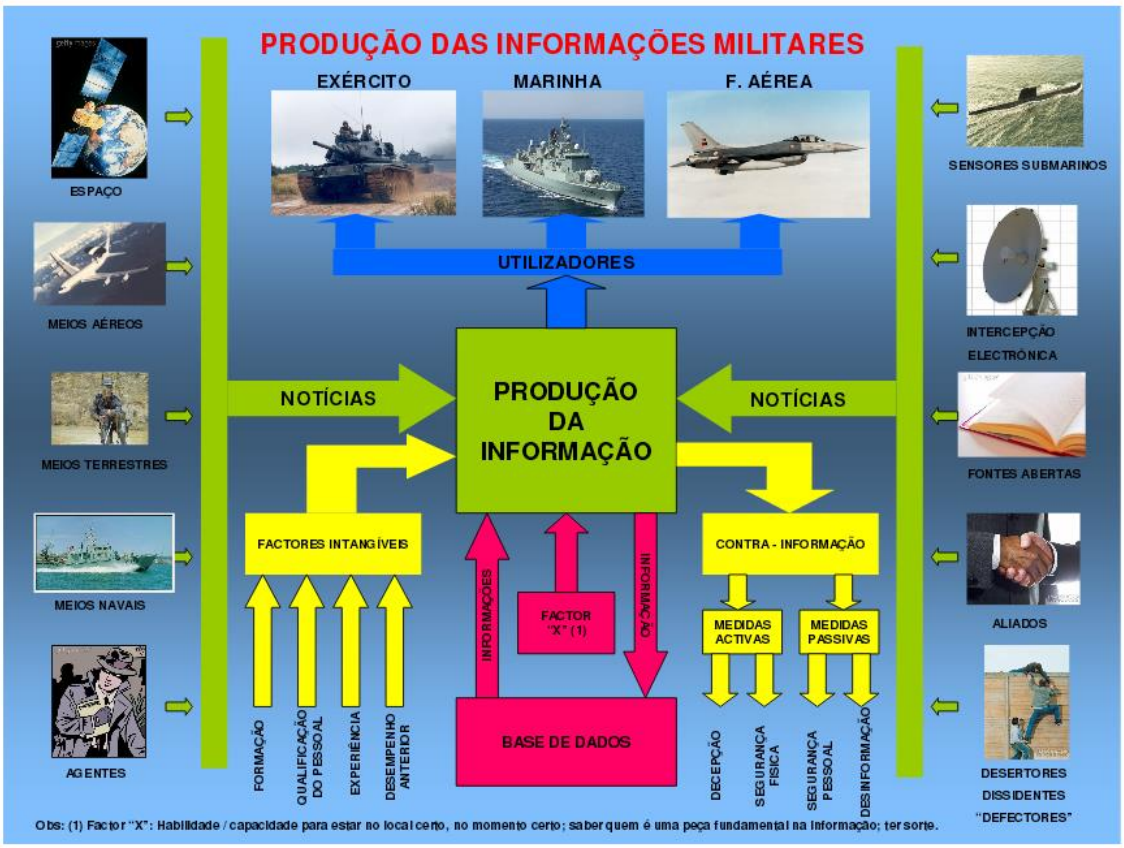


Figura 1-Produção das informações militares. A conjugação dos vários ramos de intelligence

Todas as formas de intelligence pretendem ajudar na tomada de decisão. De todas elas, a OSINT tem um papel diferente e crucial entre todas: havendo fontes, poderemos detectar no presente, acontecimentos futuros. Devido à sua capacidade de colocar um contexto nas informações das diversas *intels*, é também mais simples e sem risco humano, verificar o impacto real que as acções tomadas tiveram (ou não) na sociedade.
Também se verifica que embora todas as formas tenham um objectivo bem definido, é essencial a utilização de todas, para ter uma informação completa e uma decisão acertada, por exemplo numa guerra.

2.5 Fontes abertas – Sua definição e importância

"...tem toda a importância definir o que se deve entender por 'fontes abertas', uma vez que toda a informação assim recolhida, será passível de utilização legal e por essa via, de utilização em Tribunal, uma vez que em nenhum momento constitui uma violação da privacidade..."

Rogério Bravo[17]

A OSINT trabalha apenas com informações obtidas através de *intelligence*/informações abertas, ou seja, não classificadas. Importa então distinguir o que são fontes abertas e fontes fechadas.

Fontes abertas podem são jornais, revistas, newsletters, blogs, televisão, rádio, documentos académicos, entre outros. Estima-se que cerca de 80%[18] dos dados que precisamos para um determinado assunto, estejam disponíveis em fontes abertas. Um site/blogue/fórum pode ser uma fonte aberta mesmo que para aceder à informação se tenha de proceder a uma autenticação com palavra-passe ou outra.

Informações obtidas na *Dark Web*, ou que não estejam indexadas, assim como ficheiros obtidos por qualquer via, por exemplo, *FTP*[19] também são OSINT.

Fontes fechadas como por exemplo, documentos classificados, ou documentos obtidos por vias de espionagem, furto/roubo ou outra via invasiva e/ou não-pública, não são OSINT. A este respeito, R.Bravo, refere[20] em "Open Sources na investigação do cibercrime: conceitos e implicações" que "*a caracterização de uma fonte como "fechada" fá-la ficar fora dos limites legais de recolha de informação por iniciativa própria da policia. Pretende-se com isso, manter a privacidade do cidadão, evitar a inadmissibilidade da utilização da informação assim obtida na investigação criminal e, no limite, evitar que sobre o agente policial ou a Organização para a qual trabalha, possa recair responsabilidade civil, disciplinar e criminal".*

Se um documento classificado estiver disponível publicamente, em teoria, esse documento pode ser utilizado como OSINT.

A prática de obtenção de informações via OSINT inclui uma grande variedade de fontes de informação e que estão disponíveis publicamente:

- Meios de comunicação: jornais, revistas, rádio, televisão, e informações com base na Internet/computador
- Comunidades baseadas na Internet: desde os *blogs,* ao conteúdo gerado pelos utilizadores quer seja em sites de redes sociais, sites de partilha de ficheiros e/ou vídeos, *wikis*, bibliotecas, *blogs* e *folksonomies*.

[17]Rogério Bravo, inspector chefe da PJ, em "O conceito de Fontes Abertas na Investigação do Ciber Crime", 2014, disponível *online* em http://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime

[18] segundo o infosecInstitute, 90% segundo o RecordedFuture.com 80%

[19] FTP – file transfer protocol

[20] Open Sources na investigação do cibercrime: conceitos e implicações, disponível online em https://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime

- Dados governamentais e/ou públicos: relatórios do governo, orçamentos, aberturas de concursos, demografia, discursos, sites oficiais, avisos de segurança, adjudicações de contratos, conferências de imprensa;
- Observação de aviões amadores, monitores de rádios, satélites, fotografias;
- Fotografias de alta resolução e mapas na internet, muitas vezes comentadas por visitantes, e até cartas cartográficas de qualidade militar, muito difíceis de obter não fosse a Internet (e partilhas por seus utilizadores);
- Profissionais e académicos: conferências, simpósios, associações profissionais, trabalhos académicos e especialistas no assunto;

Note-se que não estão nem podem ser aqui incluídas as fontes de espionagem ou que recorrem a este tipo de meios ou que de alguma furem algum mecanismo técnico de segurança.

A vantagem das fontes abertas sobre fontes fechadas, está ao nível da capacidade de aquisição de informação sem criação de conflitos ou utilização de comportamentos e/ou acções ilegais. Com a Internet e a multiplicidade de fontes, qualquer pessoa consegue obter informações sobre praticamente tudo.

No site da americana CIA, é bem patente a utilidade da OSINT[21], para analisar por exemplo, se o discurso do presidente foi bem aceite por outros países:

"open sources can tell us how various groups overseas react to a speech by the president," …,"We don't have to settle for the 'official' view but can assess various groups' perceptions as well as track trends over time."

Também nas forças de segurança e de investigação criminal, se reconhece a importância e potencialidades das fontes abertas como uma forma de combater o crime o cibercrime. Segundo Rogério Bravo[22] em "*OPEN SOURCES IN CYBERCRIME INVESTIGATION* "[23], *a recolha de inteligência nas fontes abertas é uma importante ferramenta para as polícias, embora se reconheça um vazio legal nas definições dos sistemas criminais europeus o que restringe e em certos casos, até leva a abandonar importantes fontes de informação. Urge, portanto, caracterizar formalmente as fontes, pois uma fonte dita "fechada", «fá-la ficar fora dos limites legais de recolha de informação por iniciativa própria da polícia». Pelo contrário, uma fonte aberta pode ser usada em tribunal. Ainda no mesmo paper, podemos ler que no artigo 25 do ECD «... a Europol pode recolher e coligir informação, incluindo dados pessoais, a partir de fontes publicamente disponíveis.»[24]".*

Isto significa por exemplo que os órgãos de investigação criminal podem procurar e obter informação, incluindo dados pessoais, se os mesmos estiverem em fontes disponíveis na Internet (e é sabido que a Internet tem um imenso potencial já que são as próprias pessoas que lá colocaram voluntariamente, e de diversas formas, informação, prescindindo a maioria das vezes da sua privacidade.).

---

[21]INTellingence: Open Source Intelligence, disponível online em https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html

[22]Inspector-chefe da Polícia Judiciária (Portugal) e autor de algumas publicações nesta área

[23] "OPEN SOURCES IN CYBERCRIME INVESTIGATION: concept and implications" disponível online em http://www.academia.edu/7301278/OPEN_SOURCES_IN_CYBERCRIME_INVESTIGATION_concept_and_implications

[24]ECD - ACTS ADOPTED UNDER TITLE VI OF THE EU TREATY, disponível em https://www.europol.europa.eu/sites/default/files/council_decision.pdf; acedido a 2016/02/08; "Artigo 25: Information from private parties and private persons"

Como confirmar se uma fonte é aberta? Se a fonte está por defeito, acessível a qualquer pessoa (ainda que tenha de se registar/autenticar). Não pode, no entanto, contornar qualquer barreira técnica de protecção (não pode ser alvo de ataque informático por exemplo). RBRAVO[25] refere relativamente a estas potenciais dúvidas:
"…se a conduta praticada não é expressamente proibida pelo direito processual penal de um País; - se a conduta praticada é proporcional ao objectivo pretendido, ou seja, necessária, adequada e proporcional, e em simultâneo garantir o mínimo dano aos direitos, liberdades e garantias dos cidadãos; - se a conduta praticada não dependeu da ultrapassagem e ou da anulação de alguma forma de protecção de carácter técnico do serviço ou plataforma em causa…".

2.6 OSINT - Vantagens e comparação

As vantagens da OSINT são óbvias. As desvantagens nem por isso. É necessário definir claramente na Lei Portuguesa o que são e não são fontes abertas é útil para as polícias e órgãos de investigação criminal. Também para os serviços de informações, definir o que é um e outro, pode fazer a diferença. Tal como o que são dados e sua distinção de meta-dados.

Custo. Uma das maiores vantagens na utilização do OSINT é o custo muito inferior, quando comparado com as formas tradicionais de colecta de informação (este custo envolve recursos humanos, materiais e financeiros). Também no uso por estudantes, as fontes abertas estão mais disponíveis e são mais facilmente defendidas. As fontes académicas também são OSINT.

Acesso. Por estarem disponíveis em todo o lado e de forma pública:

- podem ser recolhidas, verificadas, comparadas;
- podem ser partilhadas com terceiros;
- não oferecem problemas legais;
- e mesmo através de autenticação de utilizador e/ou uso de passes, o seu acesso é permitido desde que não se contornem barreiras e/ou protecções técnicas (não recorrer por exemplo a *hacking*);

  Para quase todos os fins, seja de segurança, seja académico, o uso de fontes fechadas não é admissível.

Segurança e Privacidade. Por serem de domínio público, as informações recolhidas não oferecem ameaça ao seu possuidor (ao contrário das informações classificadas) nem usam de meios/métodos que possam por em causa a privacidade de terceiros. Em termos de segurança informática possibilita, mas não se limitando, a:

[25] "OPEN SOURCES IN CYBERCRIME INVESTIGATION: concept and implications" disponível online em http://www.academia.edu/7301278/OPEN_SOURCES_IN_CYBERCRIME_INVESTIGATION_concept_and_implications e https://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime

- *Cyber Threat Intelligence*, obter informações, processá-las e gerar "*feeds*" que podem ser integrados em sistemas *SIEM*, e ser utilizadas por equipas de resposta a incidentes informáticos
- Segurança de organizações, com o estudo do *background* dos funcionários e seu contentamento actual com a empresa e seus funcionários
- *Competitive Intelligence,* o que faz o nosso concorrente? Produtos e/ou ideias melhores? Será que estamos a colocar demasiada informação nos nossos próprios sítios web?

2.7 OSINT - Utilização pela Defesa e sociedade civil

Forças de segurança e órgãos de investigação portugueses: redes sociais em que todos falam de tudo, inclusive, em que se preparam manifestações e ataques informáticos. Hoje, é fácil fazer queixa online. Já não é necessário dirigirmo-nos a uma esquadra. Em termos de investigação, podemos procurar online por x pessoa, saber o seu background, conhecimentos, etc, útil por exemplo se a quisermos contratar ou se esta gerar suspeitas, localizar pessoas desconhecidas, … Algumas fotos que se colocam online não são devidamente "anonimizadas" e os metadados podem revelar bastante mais do que o modelo da máquina que tirou a foto.

A GNR investiu 754.000,000€[26] [27] [28] [29] [30]na aquisição de software OSINT(ver "Órgãos de investigação criminal e forças de segurança"), o que demonstra a abertura e a consciência das forças de segurança para este fenómeno.

A PSP embora sem um centro OSINT, também dá importância a esta forma de intelligence, veja-se por exemplo, o *powerpoint* interno,"actualização de informações OSINT e HUMINT"[31], em que é mostrado o plano da PSP para protecção do jogo Portugal vs Suécia e a sua colaboração com as entidades suecas.

Informações sempre actualizadas. Estas informações estão constantemente a fluir, a ser actualizadas e a remeter para locais e acontecimentos, quase em tempo real. Com a Internet qualquer tópico está *online* e tem elementos de interesse (imagem, som, texto, possivelmente até vídeo). Se não fora alguns vídeos colocados no *Youtube*, algumas situações não tinham sido criminalmente punidas.

---

[26] www.gnr.pt/ficheiros/seguranca_interna/3.pdf

[27] https://www.sg.mai.gov.pt/Noticias/Paginas/Fundo-de-Seguran%C3%A7a-Interna-financia-Unidade-OSINT-da-Guarda-Nacional-Republicana.aspx

[28] Contrato Público, Aquisição de Software de Análise Ibm I2 Para O Centro de Informações/Osint da Gnr https://www.racius.com/aquisicao-de-software-de-analise-ibm-i2-para-o-centro-de-informacoes-osint-da-gnr/

[29] Documento interno de adjudicação, assinado e digitalizado, disponível online em http://www.base.gov.pt/base2/rest/documentos/319221

[30] Caderno de encargos da aquisição da gnr, http://www.gnr.pt/mostrarPdf.ashx?a=8&i=162&download=0&f=5

[31] http://app.parlamento.pt/webutils/docs/doc.pdf?path=6148523063446f764c324679626d56304c334e706447567a4c31684a5355786c5a793944543030764f454e4651304d76523152454c305276593356745a57353062334e4259335270646d6c6b5957526c5132397461584e7a59573876597a49334d7a52695a5749745957a4d5330305a4451334c54686a4d7a45744e5755774e6d4a6b5a446c6b4d3255784c6e426b5a673d3d&fich=c2734beb-ab31-4d47-8c31-5e06bdd9d3e1.pdf&Inline=true

Universalidade. Houve, há e haverá (cada vez mais, assim haja liberdade de expressão) OSINT em tudo, em todos e em todo o lado.

Vantagens académicas. Por hoje existirem imensas formas de partilha de informação, sobre tudo, e qualquer assunto, é possível estudar qualquer cultura, esteja em que língua e local for. É possível estudar, aprender e adaptar as nossas forças de segurança e Defesa, a potenciais ameaças. Algumas coisas poderão ser pagas, mas estão acessíveis a todos.

Forças de Segurança e Defesa. Mapas partilhados por qualquer pessoa, com especial relevo para o Google Maps e seus incríveis mapas, assim como a possibilidade de percorrer ruas. Meta-dados em imagens podem indicar onde e quando, é que foi tirada a foto ao míssil. Preparando não só as pessoas, mas também os meios.

As potencialidades da OSINT são muitas, mas num breve resumo, é possível hoje fazer na Internet, pesquisas que devolvem em segundos o que antes precisava de meses. E nem assim se obtinha tanta informação. Além dos riscos serem elevados: levantar suspeitas, ser acusado de intrusão na vida alheia, espionagem, … A a maior parte das vezes e da informação nem seria possível de obter. Potenciais informações obtidas *online* através de OSINT (ainda que pareçam muito óbvias):

- Dados pessoais (incluindo orientações sexuais) que ninguém na própria rua sabe, mas que estão expostos nas redes sociais, quem são os amigos, onde trabalha, trabalhou e onde estudou, amigos de profissão, qualificações académicas, …
- Nomes completos
- Endereços de email
- Números de telefone
- Fotografias da pessoa, família, animais, automóveis, rua, ...
- Localização de barcos, aviões, pessoas, …
- Informações sobre sistema operativos, navegador *web*, software instalado, …
- Informações do IP, alojamento, DNS, …
- Geo-localização através de metadados, através do endereço IP e metadados
- Recuperação de informações de sítios web que hoje em dia já nem existem ou foram apagados
- Compras, gostos(*likes*)
- Mapas de 2D e 3D de uma infinidade de sítios

Todas as informações acima listadas são úteis, e são aproveitadas por pessoas com intenção maliciosa (ladrões, cyber-*stalkers,* terroristas, …), policias, detectives privados, agências de informação, grupos de informações noticiosas, …

2.8 OSINT - Desvantagens e limitações à recolha e/ou tratamento

"Just because open source is 'free' or publicly available doesn't mean it is easy"

CIA[32]

Existem mesmo desvantagens na obtenção de OSINT? Sim. Os limites da obtenção de OSINT são maiores do que inicialmente pensado, e no decorrer desta dissertação foram aparecendo algumas pensadas, outras por vozes de velhos do Restelo, outras que legalmente também foram colocadas, outras que no decorrer do tempo e segundo novas ideias vindas com as novas leis europeias da privacidade e protecção dos dados pessoais (RGPD[33]), podem vir a colocar "entraves" e vir a "indefinir" a noção de OSINT neste espaço. Seguem a seguir algumas das desvantagens identificadas.

Segundo o sítio Web do FBI[34], ao contrário de outras formas de inteligência, a OSINT não tem uma responsabilidade de aquisição única (*tem múltiplas formas de aquisição e muitas fontes*). Uma grande vantagem é o acesso por qualquer um em qualquer lugar. No entanto, a sua aquisição nos dias de hoje, gera tão grande quantidade de informação que se pode tornar difícil encontrar informações de valor e de confiança. É por isso importante ser analisada/filtrada por analistas para lhe ser dado uso posterior. As desvantagens podem ser sumariadas nas seguintes:

Multiculturalidade e conhecimentos: para analisar OSINT de um site português de notícias locais, é simples (em teoria). Para analisar OSINT de um site Vietnamita, já não será assim, tal como não será, se o OSINT vier de um país com uma cultura diferente, com políticas diferentes, etc. É necessário filtrar, perceber e analisar muita informação que sai 24h por dia, todos os dias, e isso obriga a ter pessoas e recursos materiais, assim como perceber a origem e a realidade da fonte da notícia, para a classificar ou descartar. *"The ability to combine foreign language skill, cultural knowledge, and advanced search techniques is not common."*

Clara definição de fontes abertas. Como anteriormente referido (Vantagens e Desvantagens), é necessário estar bem definido o que são fontes-abertas. De outro modo, poder-se-á comprometer investigações de Polícia e/ou prejudicar culpados e/ou inocentes.

Concorrência e *competitive intelligence*. Vantagem ou desvantagem? Depende de quem usa. É claramente uma vantagem visto tratar-se de uma forma de inteligência passiva, sem impacto visível, mas que permite estudar o adversário económico, entender o que faz, como faz, e depois aprender os pontos fortes, atacando os fracos. Sendo OSINT é possível verificar por exemplo, os preços de um hotel na Internet e baixar os nossos. Uma forma que não é OSINT e é bastante grave, é a espionagem industrial, que tenta roubar segredos de produção e/ou económicos. Com o intuito de proteger esta propriedade intelectual e segredos de produção, o Serviço de Informações de Segurança Português, tinha (à data da escrita não se

[32] https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html

[33] RGPD – Regulamento Geral de Protecção de Dados, disponível online no site europeu de legislação, *RÈGLEMENT (UE) 2016/679 DU PARLEMENT EUROPÉEN ET DU CONSEIL* em https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679. Consultado também: https://protecao-dados.pt/o-regulamento/

[34] *Site* do FBI, disponível online em https://www.fbi.gov/about-us/intelligence/disciplines

encontrou) uma zona na sua página *web* para consciencializar deste perigo e oferecer apoio às empresas e instituições.

Limites à recolha de OSINT: recentemente, devido à introdução da nova Lei europeia RGPD, que define um conjunto de directrizes que são necessárias cumprir[35], foram colocadas algumas limitações (*confusões*) extra quer à recolha quer ao tratamento de dados OSINT. No geral é permitida a recolha de OSINT, se não for proibida (expressamente) pelas leis criminais de um Estado, ou se o comportamento/conduta praticada para recolher também não seja expressamente proibida (espionagem por exemplo não é permitido legalmente, não trabalha com fontes abertas por isso não é OSINT). Se a recolha de informação for de fontes abertas mas tiver de ser trabalhada e as conclusões saírem para domínio público, será necessário ter alguma forma de autorização dos visados (embora a informação original seja OSINT, a conclusão que tiramos e difundimos pode de alguma forma violar a privacidade e os direitos dos visados, por exemplo, poderíamos ser ameaçados por difamação ou violação da privacidade, entre outros).

Como anteriormente referido, a obtenção de informação não pode envolver a eliminação ou contorno de qualquer obstáculo seja ele físico (uma porta fechada) ou lógico (uma qualquer forma de autenticação que peça uma passe e nós não a tenhamos). As informações OSINT que se obtenham quebrando passes para entrar numa zona reservada de um sítio web não são válidas e é crime. Por outro lado, se o sítio web exigir uma passe e nós tenhamos de nos registar, e assim obter informações que qualquer pessoa poderia obter, isso é OSINT.

A própria policia tem limites quanto à sua forma de adquirir informação, e uma fonte fechada não pode ser utilizada. Desta forma, protege-se a privacidade do cidadão e limitam-se potenciais abusos. Esta limitação protege os direitos e liberdades do cidadão comum, mas também a própria policia que assim não pode ser acusada de ter arranjado informações de forma ilegal, nem incorre em procedimentos civis, criminais ou de carácter disciplinar.

No capítulo “Avaliação”, teremos as limitações OSINT encontradas no decurso deste trabalho.

2.9 Espionagem

2.9.1 Agentes, espiões e serviços de informações

O crime de espionagem (seja a tradicional ou a a ciber), é dos crimes mais graves, sendo punível em alguns países com a própria morte. Funciona com fontes fechadas e, portanto, não pode ser OSINT.

Os elementos a actuar nos Serviços de informações (e que poderão nem sequer fazer parte daquela estrutura, mas trabalhar/colaborar com/para ela), são designados, se forem nacionais, por agentes, e se forem estrangeiros, por espiões. O seu método assenta principalmente no disfarce e intrusão junto do inimigo, com o objectivo de saber informações

[35] RGPD – Regulamentação e coimas https://protecao-dados.pt/o-regulamento/ e documento completo em formato digital PDF, https://protecao-dados.pt/wp-content/uploads/2017/07/Regulamento-Geral-Prote%C3%A7%C3%A3o-Dados.pdf

do seu *modus operandi*, objectivos, formas de financiamento, alvos, inimigos, chefias, entre outros.

Em Portugal, *a espionagem é um crime punido por lei*[36]. No site deste serviço[37], *é dito "O Serviço de Informações de Segurança (SIS) é o único organismo incumbido da produção de informações que contribuam para a salvaguarda da segurança interna e a prevenção da sabotagem, do terrorismo, a espionagem e a prática de actos que, pela sua natureza, possam alterar ou destruir o Estado de direito constitucionalmente estabelecido."*

Os agentes e os espiões respondem a ordens e prioridades, pedidas pelos seus serviços/governos e que envolvem o acesso privilegiado e em "primeira mão", a informações de natureza classificada (secreta, confidencial, …), sempre sensíveis e que está relacionada com as áreas:

- Económica – informações de teor tecnológico, industrial e comercial, de empresas e instituições públicas, que permitam ganhar vantagem económica (ou não sofrer), por parte das empresas e instituições nacionais. Nacionalmente, em 2006 o SIS criou[38] "…*o Programa de Segurança Económica, dirigido para empresas e para organismos do Estado. Neste domínio, prevalece a capacidade de antecipação das ameaças e o conhecimento do modo de atuar dos seus agentes, colocados ao serviço da salvaguarda do bem-estar económico e social da nossa sociedade."*
- Militar – informações relacionadas com missões, organizações militares de Defesa (forças armadas) e Segurança nacionais, vulnerabilidades e capacidades (tamanho, organização, quantidade, …), assim como as movimentações e exercícios. Aliados políticos podem ter impacto na Defesa.
- Política – informações de teor politico quer interno, quer externo. Saber qual a posição dos Governos em relação ao nosso. Diz também o site do SIS que "*Alguns governos estrangeiros, através dos seus serviços de informações, procuram ainda exercer controlo sobre as comunidades residentes no exterior, tanto para limitar o seu exercício pleno da cidadania, como para fins de espionagem."*

Segundo o SIS[39], "*a espionagem consiste na obtenção de informação que, pelo seu valor e relevância para o interesse nacional, está protegida por medidas de segurança. O acesso ilícito a essa informação faz-se através de métodos clandestinos, recorrendo a meios técnicos cada vez mais sofisticados ou a agentes e fontes humanas que se encontram ao serviço dos interesses políticos, militares e económicos de um Estado estrangeiro".*

O site da *wikipedia* (*vale o que vale*) diz-nos[40] o seguinte acerca disto que "…*Nenhum serviço secreto de Estado usa a palavra "espionagem" no seu nome ou para descrever sua atividade de colheita de informações ou inteligência, embora todos declarem fazer contra-espionagem. Muitas nações espiam rotineiramente seus inimigos, mas também seus aliados, embora sempre o neguem. A duplicidade que envolve a utilização do termo espionagem deve-se ao facto de essa atividade ser frequentemente ditada por objectivos secretos e interesses*

[36] *Artigo 317º, do Código Penal) e, nos termos do Artigo 3º da Lei Orgânica nº 9/2007, 19 de Fevereiro*
[37] https://www.sis.pt/ameacas
[38] https://www.sis.pt/ameacas - contra-espionagem
[39] SIS - Definição de espionagem - https://www.sis.pt/ameacas
[40] Disponível online em https://pt.*wikipedia*.org/wiki/Espionagem, revisitado em 2017-12-21

inconfessáveis publicamente, enquanto que nos rivais ou inimigos ela é sempre denunciada e condenada..."

Os serviços de informações não utilizam apenas HUMINT, mas sim, todos os ramos de inteligência que existem, se isso lhes for favorável para a conclusão da missão.

O disfarce pode pasar pela visita a instalações com o propósito de visita, ou de oferecer um serviço. Pode também passar pela promoção de projectos comuns, eventos públicos de sensibilização e propaganda.

Como referido anteriormente, diz a própria CIA[41] que não precisa de ser informação classificada para ser de elevada importância: "***Information does not have to be secret to be valuable**. Whether in the blogs we browse, the broadcasts we watch, or the specialized journals we read, there is an endless supply of information that contributes to our understanding of the world."*

2.9.2 Combate à (ciber)espionagem

Para combater a espionagem e a ciberespionagem, assim, como o cibercrime e outros que recorram à criminalidade por via tecnológica, foi criada em 2016[42], na Polícia Judiciária, a Unidade Nacional de Combate ao Cibercrime e à Criminalidade Tecnológica. Esta nova unidade pretende ser uma *"resposta estrutural, preventiva e repressiva ao fenómeno do cibercrime e do ciberterrorismo, e que é inspirada no modelo adotado pelo EC3 (European Cybercrime Center) da EUROPOL, cujos pontos focais são o abuso sexual de crianças através da Internet, a fraude com os cartões e outros meios de pagamento eletrónico e virtuais, a criminalidade informática pura (os crimes previstos na Lei n.º 109/2009, de 15 de setembro) e a criminalidade praticada com recurso a meios informáticos."*

Para os objectivos desta dissertação interessam-nos especialmente as competências desta nova unidade, designada por UNC3T no âmbito da espionagem: "*v) De espionagem, quando cometido na forma de um qualquer programa informático concebido para executar ações nocivas que constituam uma ameaça avançada e permanente."*

Também nacionalmente, foi criado em 9 de Maio de 2014, o Centro Nacional de CiberSegurança[43] (embora inicialmente previsto já desde 2012), que veio também a ter responsabilidades na luta à espionagem, contra-ciberterrorismo, etc. Um dos objectivos da sua criação, era "*contribuir para que o país use o ciberespaço de uma forma livre, confiável e segura, através da promoção da melhoria contínua da cibersegurança nacional e da cooperação internacional, em articulação com todas as autoridades competentes, bem como da implementação das medidas e instrumentos necessários à antecipação, à deteção, reação e recuperação de situações que, face à iminência ou ocorrência de incidentes ou ciberataques,*

[41] Disponível online em https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html, consultado 28/08/2018

[42] Decreto-Lei n.º 81/2016 - Diário da República n.º 228/2016, Série I de 2016-11-28

[43] Publicação da criação do CNCS (Centro Nacional de CiberSegurança) em Diário da República, 1.ª série — N.º 89 — 9 de maio de 2014

ponham em causa o funcionamento das infraestruturas críticas e os interesses nacionais."

Entre várias das suas competências[44], é referido que o CNCSeg "*atua ainda em articulação e estreita cooperação com as estruturas nacionais responsáveis pela ciberespionagem, ciberdefesa, cibercrime e ciberterrorismo, devendo comunicar à Polícia Judiciária, no mais curto prazo, os factos de que tenha conhecimento relativos à preparação e execução de crimes… ".*

De entre várias técnicas de combater a espionagem, em que a prevenção é uma delas[45], existe ainda a contra-espionagem, que "*visa prevenir e impedir as atividades de um governo, associação, organização ou serviço de informações estrangeiros, ou de um agente seu, que impliquem prejuízo para o interesse nacional".* A contra-espionagem é uma das funções do SIS, que para esta função, *recolhe, analisa e difunde informações tendentes à neutralização dos agentes que promovem atividades de recolha ilegal de informação com valor estratégico para o Estado".*

Também com o objectivo de combater a espionagem (*ciber* incluída), o SIS criou em 2006, o Programa de Segurança Económica, dirigido para empresas e para organismos do Estado. Aposta-se na capacidade de antecipação das ameaças e no capacitar das pessoas e indústria de identificar e conhecer o modo de atuar dos espiões (quer nacionais competidores quer internacionais de governos e empresas estrangeiras). O Combate à espionagem económica previne que segredos industriais sejam furtados e copiados, ajudando a proteger assim os interesses do cidadão comum, do Estado e da economia nacional.

Como já referido, também são competências da Polícia Judiciária, Centro Nacional de CiberSegurança, mas não só. Todos podem ajudar. O iKNOW nasce também com esse objectivo.

2.9.3 Lei Portuguesa para espionagem

Segue-se alguma legislação seleccionada e actual sobre o assunto, que contém referências à luz da Lei, sobre as possibilidades e limitações dos serviços de informações portugueses no âmbito da luta à espionagem, intercalada com alguma informação sobre *fake news* e outros:

Lei Orgânica n.º 4/2017[46]
Sumário: aprova e regula o procedimento especial de acesso a dados de telecomunicações e Internet pelos oficiais de informações do Serviço de Informações de Segurança e do Serviço de Informações Estratégicas de Defesa e procede à segunda alteração à Lei n.º 62/2013, de 26 de agosto (Lei da Organização do Sistema Judiciário)

[44]Decreto-Lei n.º 69/2014, Presidência do Conselho de Ministros, disponível *online* em http://dre.pt/util/getpdf.asp?s=diad&serie=1&iddr=2014.89&iddip=20140696

[45] através de programas que se destinam a sensibilizar para os riscos desta ameaça e a estimular comportamentos de segurança adequados

[46] Diário da República n.º 164/2017, Série I de 2017-08-25, disponível online em https://dre.pt/home/-/dre/108052020/details/maximized

A Assembleia da República decreta, nos termos da alínea c) do artigo 161.º da Constituição, a lei orgânica seguinte:

Artigo 1.º
1 - A presente lei regula o procedimento especial de acesso a dados previamente armazenados pelos prestadores de serviços de comunicações eletrónicas que se mostrem estritamente necessários para a prossecução da atividade de produção de informações pelo Sistema de Informações da República Portuguesa (SIRP) relacionadas com a segurança interna, a defesa, a segurança do Estado e a prevenção da espionagem e do terrorismo, o qual é sujeito a acompanhamento do Ministério Público e controlo judicial.

Artigo 3.º - Acesso a dados de base e de localização de equipamento
Os oficiais de informações do SIS e do SIED podem ter acesso a dados de base e de localização de equipamento para efeitos de produção de informações necessárias à salvaguarda da defesa nacional, da segurança interna e da prevenção de atos de sabotagem, espionagem, terrorismo, proliferação de armas de destruição maciça e criminalidade altamente organizada e no seu exclusivo âmbito.

Artigo 4.º - Acesso a dados de tráfego
Os oficiais de informações do SIS e do SIED apenas podem ter acesso a dados de tráfego para efeitos de produção de informações necessárias à prevenção de atos de espionagem e do terrorismo.

Artigo 10.º - Apreciação judicial
2 - O acesso dos oficiais de informações do SIS e do SIED a dados de tráfego só pode ser autorizado no quadro da produção de informações de prevenção da espionagem e do terrorismo.

Artigo 13.º - Factos indiciários de espionagem e terrorismo
Os dados obtidos que indiciem a prática de crimes de espionagem e terrorismo são imediatamente comunicados ao Procurador-Geral da República para os devidos efeitos.

Artigo 17.º - Alteração à Lei da Organização do Sistema Judiciário
4 - No Supremo Tribunal de Justiça há também uma formação das secções criminais, constituída pelos presidentes das secções criminais e por um juiz designado pelo Conselho Superior da Magistratura, de entre os mais antigos destas secções, que procede ao controlo e autorização prévia da obtenção de dados de telecomunicações e Internet no quadro da atividade de produção de informações em matéria de espionagem e terrorismo do Serviço de Informações de Segurança e do Serviço de Informações Estratégicas de Defesa.
A Lei Portuguesa tem no Artigo 317.º - Espionagem[47], pesadas medidas para quem for julgado culpado por este crime. Diz o mesmo que é acusado e punido com pena de prisão de 3 a 10 anos quem:

[47]Código Penal Português, disponível online em http://www.codigopenal.pt/index5.html

a) *Colaborar com governo, associação, organização ou serviço de informações estrangeiros, ou com agente seu, com intenção de praticar facto referido no artigo anterior; ou*

b) *Recrutar, acolher ou receber agente que pratique facto referido no artigo anterior ou na alínea anterior, ou, de qualquer modo, favorecer a prática de tal facto;*

É punido com pena de prisão de 5 a 15 anos, *"Se o agente praticar facto descrito no número anterior violando dever especificamente imposto pelo estatuto da sua função ou serviço, ou da missão que lhe foi conferida por autoridade competente…"*

Já o Artigo 330.º - Incitamento à desobediência colectiva, é de especial relevância se tivermos em conta a produção actual e sem limites de contra-informação e *fake news* que se vem vindo a assistir nas redes sociais, *blogs* e órgãos de informação (que sem o saber, podem ajudar na difusão destas notícias). Diz o artigo *330.º - Incitamento à desobediência colectiva:*

- 1 - Quem, com intenção de destruir, alterar ou subverter pela violência o Estado de direito constitucionalmente estabelecido, incitar, em reunião pública ou por qualquer meio de comunicação com o público, à desobediência colectiva de leis de ordem pública, é punido com pena de prisão até 2 anos ou com pena de multa até 240 dias.
- 2 - Na mesma pena incorre quem, com a intenção referida no número anterior, publicamente ou por qualquer meio de comunicação com o público:
- Divulgar notícias falsas ou tendenciosas susceptíveis de provocar alarme ou inquietação na população;
- Provocar ou tentar provocar, pelos meios referidos na alínea anterior, divisões no seio das Forças Armadas, entre estas e as forças militarizadas ou de segurança, ou entre qualquer destas e os órgãos de soberania;
- ou incitar à luta política pela violência.

O Artigo 331.º, "Ligações com o estrangeiro" traz também a ideia de espionagem, *humint* e *fake news*, embora se refira ao estrangeiro e seus agentes, como "ligações com o estrangeiro. Com pena de prisão até 5 anos, "se pena mais grave lhe não couber por força de outra disposição legal", diz o mesmo:

Quem, com intenção de destruir, alterar ou subverter pela violência o Estado de direito constitucionalmente estabelecido, se puser em ligação com governo de Estado estrangeiro, com partido, associação, instituição ou grupo estrangeiro ou com algum dos seus agentes para:

- Receber instruções, directivas, dinheiro ou valores; ou
- Colaborar em actividades consistindo:
    - Na recolha, preparação ou divulgação pública de notícias falsas ou grosseiramente deformadas;
    - No aliciamento de agentes ou em facilitar aquelas actividades, fornecendo local para reuniões, subsidiando-as ou fazendo a sua propaganda;
    - Em promessas ou dádivas;
    - Em ameaçar outra pessoa ou utilizar fraude contra ela;

A este respeito, foi recentemente notícia[48], um agente(espião?) do SIS que foi "apanhado" a alegadamente estar a colaborar com órgão de informações estrangeiro e a vender informações nacionais e militares NATO. *"Dentro dos crimes de que é acusado Frederico Carvalhão, o de espionagem é o que tem uma pena mais grave com punições entre 5 e 15 anos de prisão. Quem violar o "dever especificamente imposto pelo estatuto da sua função ou serviço, ou da missão que lhe foi conferida por autoridade competente" enfrenta uma punição especialmente mais severa, como é o caso do espião português."*

2.9.4 Insiders

Insiders[49]. Os espiões podem ser qualquer pessoa, estar em qualquer sítio a qualquer hora. O site *State of Security[50]* referia-nos que a maior ameaça para 2017, eram os ataques feitos por *insiders[51]*, que teoricamente deviam ser as que mais defendiam a própria casa. Eis porquê:

- Temporalidade - pode permanecer indetectável por anos ou para sempre. Quanto mais tarde for, pior podem ser os custos para a organização, podendo até mesmo ser o seu fim.
- Dificuldade de detecção – Dificuldade de distinguir acções normais de trabalho ou descuido, de acções de sabotagem. Dificuldade de saber as reais intenções do trabalhador quando lida com informações ou quando age.
- Facilidade de cobrir pistas – Além de ser difícil de detectar ou adivinhar intenções, é fácil para o empregado, disfarçar as suas acções. Seja no apagar de registos informáticos, seja na sabotagem deliberada seja nos descuidos ou no prejudicar outras pessoas pelos seus actos.
- Dificuldade de prova – Trabalhando no local de trabalho, tem oportunidade e justificação. Não se pode adivinhar os objectivos/intenções, e às vezes pode ser difícil atribuir a culpa de algo que aconteceu "sem querer", ou por erro do sistema, ou por qualquer erro do material, ou causas imputadas a terceiros...

O objectivo do indivíduo que envia informações para o exterior da organização pode ter diferentes motivações, mas geralmente ocorre porque não existe uma forte política de segurança implementada. Como por exemplo:

- Excesso de privilégios *e Need to know* - acesso à informação consoante a real necessidade de saber para poder trabalhar
- Sub-contratados – a empresa sub-contrata pessoal para resolução de problemas temporários. Os novos funcionários têm acesso a informação e material. Tomam conhecimento do funcionamento da empresa, localização, passes e podem inclusive

[48]Disponível online em http://www.sabado.pt/portugal/seguranca/detalhe/o-que-se-sabe-sobre-o-espiao-portugues-detido-em-roma?ref=DET_relacionadas_seguranca

[49]

[50]Insider Threats as the Main Security Threat in 2017. 11 de Abril de 2017. Disponível *online* em https://www.tripwire.com/state-of-security/security-data-protection/insider-threats-main-security-threat-2017/, revisitado em 2018-01-21

[51]Insiders - elementos internos da empresa/organização, são talvez das ameaças humanas mais dificeis de apanhar, pois têm o acesso, o conhecimento e o seu erro pode ser desculpado, não conseguindo facilmente saber se o erro(possível sabotagem) foi ou não intencional. São HUMINT.

instalar hardware e/ou software que monitorize/espie comportamentos de utilizadores e recolha dados confidenciais

- Funcionários descontentes ou despedidos – ex-funcionários e funcionários descontentes podem enviar a informação para fora, como forma de mostrar que têm razão, mostrar descontentamento ou mesmo por vingança contra alguém ou algo. Tiveram acesso a material privilegiado e se a política de empresa não for cumprida, vai levar passes e potenciais acessos remotos a sistemas e materiais que não devia ter acesso.
- Acessos a recursos indevidos – utilização de equipamentos para outros fins, utilização dos próprios equipamentos (*BYOD*) na rede informática do serviço. Comunicação nas redes sociais durante o trabalho, partilha de informações da empresa (desde fotos a amizades, passando por projectos)
- Técnicas de obtenção de informações quer por engenharia social, quer por conversas ocasionais e informais.
- Recolha de informações por via de invasão informática*/hacking, fotocópia de material, cópia de informações internas, leak* de informação para fora da empresa, …
- Vasculha de caixotes de lixo, papéis esquecidos nas impressoras.
- …

É necessário que cada empresa ou instituição faça a sua própria catalogação de informação, e faça uma análise bastante aprofundada de quem contrata, assim como a crie políticas de segurança fortes (mas simples de perceber e executar) e contacte com o seu funcionário de forma constante para saber do seu estado e atitude perante a organização (entre outros).

Os modos de actuação de um espião/insider ou outro são muito variadas e podem ser utilizadas diferentes técnicas consoante o que se pretende e o que fazem. É comum até em testes de penetração informática, mas não só, o estudo da organização que se pretende "atacar" como por exemplo:

- O estudo dos horários das empresas, seus funcionários e turnos de guardas
- O estudo dos comportamentos dos funcionários, hábitos, locais de almoço e lazer
- Engenharia social, conhecendo a empresa, seus objectivos e funcionários, é mais fácil telefonar, fingir que é um cliente ou fornecedor e pedir informações que não teria direito
- O estudo das instalações físicas, locais de acesso e saídas de emergência, para estacionar, entrar despercebido e fugas rápidas. (infiltração e/ou extracção)
- Para um funcionário é extremamente fácil circular, alegando serviço. Se for apanhado a deambular, pode dizer que se enganou ou que procurava alguém ou alguma coisa esquecida, confusão ou outro. Um funcionário pode também sabotar o trabalho de pessoas inocentes, dificultando a detecção

Relativamente aos espiões, *insiders* e sua detecção, fica claro que a sua detecção é difícil e que passa pela prevenção e monitorização, assim como pela consciencialização dos quadros de pessoal.

Salário justo, controlos biométricos nas entradas e saídas, contacto regular com o funcionário, atenção às atitudes do trabalhador e deste com os seus parceiros, correlação de dados, e um

SIEM[52] que tenha na sua configuração a data, tempo, e o local de onde este fez acessos a sistemas, podem ajudar na localização e previsão de problemas.

2.9.5 Agentes infiltrados

Agentes infiltrados, como aqueles utilizados pelas forças policiais, também são HUMINT e têm um extraordinário papel na luta contra a criminalidade organizada, mas devido à sua natureza e âmbito de actuação, são elas próprias, geradores de informações confidenciais e que não podem ser nunca públicas (não OSINT). Devido ao seu trabalho em prol do país, não é nem pode ser, consideramos, encarado como espionagem. O resultado da actuação destas fontes nunca é fornecido directamente, mas sim pela Polícia no geral, de modo a proteger estas fontes e a não alertar a criminalidade, da efectividade destes agentes. O uso destes agentes da Justiça está bem integrado na Lei, e não podem ser, consideramos, vistos como espionagem. Os objectivos dos agentes infiltrados é dar informação à Polícia para lhe dar elementos de apoio à decisão, como por exemplo: actuar já em relação a um grupo, prendendo os seus elementos? ou continuar a trabalhar, actuando depois, com vista ao desmantelamento da organização?

2.10 Juan Pujol/agente Garbo – O "espião OSINT"

Um dos mais célebres espiões (conhecidos pelo menos) da história recente, foi o espanhol Juan Pujol Garcia (nome de código Arabal/Bovril/Garbo…), que salvou milhares de vidas a militares e permitiu segundo o próprio presidente americano Eisenhower, a vitória dos aliados sobre os nazis. Resumidamente, Juan Pujol quis ajudar os aliados na luta contra os alemães. Ofereceu os seus préstimos a Espanha, que recusou, aos Ingleses, que também recusaram, e aos Alemães, garantindo-lhes que conseguiria e iria para Londres e de lá, lhes enviaria informações.

Juan Pujol mudou-se de Espanha para Lisboa, para tentar ir para Londres com um visto. Não conseguiu. Ofereceu novamente os seus préstimos a Espanha e Inglaterra, desta vez mostrando-se como agente alemão. Não quiseram. Juan, obtinha tudo o que podia sobre Londres através de fontes abertas, jornais, rádio, cartas, conversas. Inventava depois informações, baseadas em factos reais. Enviava depois estas informações inventadas para o seu contacto alemão, sem saber que estas informações também estavam a ser lidas/interceptadas pelo serviço de informações britânico MI.

Os serviços de Juan foram finalmente aceites pelo MI5, que o colocou disfarçadamente como tradutor da BBC (ver capítulo 11 – BBC). Garbo (nome de código pelo qual ficou conhecido), inventava dezenas de colaboradores numa rede fictícia de espionagem, com vida e histórias próprias, e mais tarde com factos dados pelo próprio MI5.

Juan teve como mérito, a invenção (ajudada pelo MI5) de um exército aliado que iria invadir a Europa através de Calais, tendo feito com que o exército alemão enviasse para ali a sua força maior, permitindo aos aliados entrar pelas praias da Normandia com muito menor

[52] SIEM - *security information and event management*, software que faz o correlacionamento de informações de diversas fontes(informáticas, biométricas, ...) e nos gera automaticamente alertas para possíveis problemas. São bons locais para integrar OSINT, que até ao momento não se viu.

resistência e mortes. Juan foi condecorado tanto pelos aliados como pelos alemães.

A sua eficácia residiu na recolha de fontes abertas, processamento e na sua transformação para factos inventados, mas com base verídica. Esta contra-informação não só salvou vidas como potencialmente terá evitado o envio de novos agentes alemães.

2.11 OSINT no combate ao Ciberterrorismo e cibercriminalidade

Segundo o techopedia[53], o Ciberterrorismo é definido pelo FBI[54] americano, como um ataque electrónico e premeditado contra um sistema informático, dados, aplicações ou outras informações com o único propósito de violência contra agentes nacionais ou subalternos. Pode também ter como alvo, organizações, empresas e simples pessoas individualmente. Já segundo o dicionário online, *Léxico*[55], trata-se de um movimento com motivação política com o objectivo de causar disrupção e medo na sociedade através das tecnologias da informação e computadores.

Segundo a mesma fonte os ataques podem ser classificados como:

- simples (ataque a um sistema individual)
- avançados (ataques mais sofisticados a múltiplas redes e/ou sistemas)
- complexos (ataques coordenados que podem ter um impacto de larga escala e fazer uso de ferramentas sofisticadas)

O ciberterrorismo pode ocorrer na Internet, em servidores e redes privadas e/ou governamentais. O facto de os meios electrónicos possibilitarem ser utilizados e terem impacto à distância, gera um sentimento de conforto e impunidade, que favorece o seu uso em vez de violência física, entre outros. Também em termos de custo e movimentações, é mais barato um computador do que uma arma.

A cibercriminalidade distingue-se do ciberterrorismo, devido às suas motivações e poder destrutivo, tal como na criminalidade e terrorismo convencionais.

Núria Amaral[56] na sua tese de mestrado, "*O papel dos serviços de informações no combate ao ciberterrorismo", tem uma referência a Barry Collin (1980), que já naquela altura referia que para ser ciberterrorismo, teria de se conseguir atingir um estado de ciberterror e que tal teria de ser suficientemente destrutivo ou/e perturbador para criar um tal estado de medo que fosse comparado àqueles causados pelos actos de terrorismo "tradicionais". O mesmo trabalho refere que autores como Mark Pollit (1997), consideravam que este tipo de ataques com repercussões na "vida real*" não seriam exequíveis. Mas os últimos anos têm revelado que é possível: basta recordar o caso do *StuxNet*[57] e mais recentemente alguns malwares detectados em centrais

[53] https://www.techopedia.com/definition/6712/cyberterrorism, consultado online em 03/03/2019

[54] *Federal Bureau of Investigation*

[55] https://www.lexico.com/en/definition/cyberterrorism, consultado online em 03/03/2019

[56] O papel dos serviços de informações no combate ao ciberterrorismo, o caso Português, Sandra Núria Amaral, 2014, Dissertação para o grau de Mestre, Academia Militar

[57] Stuxnet -

eléctricas nos Estados Unidos, Ucrânia, Venezuela, etc, … que provam que ataques informáticos podem ter efeitos bem reais na vida quotidiana das pessoas.

Cada vez mais dependente das novas tecnologias da comunicação, a sociedade como a conhecemos, vai estar mais à mercê deste tipo de ataques que podem efectivamente causar o terror. A maior consciencialização da OSINT permite a qualquer um ou a qualquer entidade, avaliar a sua vulnerabilidade e a sua imagem junto da comunidade.

O OSINT pode ter uma palavra na tomada de acção perante uma ameaça, detectando antes de acontecer, através da detecção atempada de palavras, vídeos, sinais, códigos.

Hackers cujo objectivo seja atacar ou espiar um Governo estrangeiro, ou incitar a uma guerra electrónica ou mesmo a utilizar *drones* com fim a provocar o pânico ou uma mensagem de medo, podem ser considerados ciberterroristas. Internamente, estes ciberterroristas podem ser empresas concorrentes que usam sabotagem electrónica para sabotar serviços ou equipamentos da concorrência, a fim de tirar o site do ar para retirar dividendos financeiros, além de uma mensagem.

Combate ao ciberterrorismo e cibercriminalidade:

Em Portugal e legalmente, a investigação e o combate à cibercriminalidade estão a cargo da Polícia Judiciária. O ciberterrorismo não tem uma entidade especifica porque pode abranger diversos domínios e portanto abranger e/ou exigir a resposta e investigação de diversas entidades.

Existem actualmente pelo mundo, diversos projectos (conhecidos pelo menos, pois acredita-se que existam bastantes mais), de entidades bastante reputadas e que pretendem ajudar as forças de segurança e Defesa a combater o ciberterrorismo e criminalidade. Estes esforços concentram-se na antevisão e detecção destes fenómenos, quer na internet quer na vida offline. Nacionalmente por exemplo, se estivermos atentos às redes sociais e algumas palavras-chave, podemos com alguma certeza, antecipar manifestações, ataques informáticos e outros problemas/ameaças que são colocadas na internet antes de acontecerem, muitas vezes como forma de atrair outros participantes. Isto foi observado aquando dos ataques nacionais do 25 de Abril por diversos grupos nacionais e internacionais que se aliaram (ou foram convidados).

Nos EUA por exemplo, a agência de projectos de Defesa DARPA[58], encontra-se a desenvolver um motor de busca para a *Dark Web*[59] que pretende ajudar as autoridades a compreender e a investigar o que se passa na *Dark Web,* para que este ambiente não seja apenas domínio de criminosos e actividades menos lícitas. Citando o artigo[60] lê-se que "*Numa das experiências foi possível detetar padrões de movimentos de pessoas relacionadas com o tráfico de humanos (traficantes e vítimas), baseando a análise no surgimento de anúncios de oferta de sexo. Esta*

[58] DARPA - *Defense Advanced Research Projects Agency*

[59] Dark web – servidores e serviços que só são alcançáveis na Internet, através de softwares, configurações ou autorizações específicas para o acesso.

[60] "DARPA está a desenvolver motor de pesquisa para a Dark Web", disponível online em http://exameinformatica.sapo.pt/notícias/internet/2015-02-12-DARPA-esta-a-desenvolver-motor-de-pesquisa-para-a-Dark-Web, 11/02/2015 11:45

experiência mostra que será possível ajudar as autoridades a detectarem crimes com maior facilidade, mesmo que sejam conduzidos através da Dark Web."

Ilegalmente e sem regras, existem na internet grupos que combatem outros grupos. Por exemplo, alguns hackers combateram os sites de aliciamento e propaganda do auto-proclamado Estado Islâmico[61]. Positivo nos fins, mas não nos meios e mesmo nos fins, poderá ter atrapalhado a acção das autoridades, Defesa e forças de segurança. No site ttps://krypt3ia.wordpress.com/ é dada a seguinte teoria:

"f you take all the sites down for however long you will only force them to make other sites that are more under the radar. You will be also teaching them about security and you don't want to be doing that do you? Say, did you see the article from Glenn Greenwald about how Iran learned from our Stuxnet attacks on them and are now a real threat? Yeah, see, it's a double edged sword kids."

Resumindo, se os *hackers* forçarem o fecho dos sites do Estado Islâmico baseando-se na indisponibilidade e outras ameaças, isto só forçaria os elementos do E.I. a encontrar outras alternativas, a aprender mais sobre segurança e a colocarem o alojamento mais escondido. O texto dá o exemplo do *Stuxnet* e de como o Irão aprendeu. Atacar nem sempre tem os resultados esperados.

Em OSINT, e neste tipo de estratégias valerá mais ir vendo o que os outros vêm sem dar nas vistas. Em caso realmente importante, age-se. Poderemos pensar nisto como a estratégia dos EUA na II Guerra Mundial relativamente a já terem a máquina Enigma funcional, mas apenas em casos extremos a utilizavam, deixando assim, afundar barcos e outras perdas consideradas "menores".

Utilizando OSINT e os dados obtidos pela *Wikipedia* relacionados com ataques terroristas foi possível criar um mapa em tempo real dos ataques terroristas pelo mundo fora, assim como associar ao grupo terrorista que lhes deu origem. O mapa pode ser acedido[62] online, com possibilidade de ser editado[63] e melhorado por todos.

[61] No site https://krypt3ia.wordpress.com/, existe a imagem lateral e o congratular dos hackers que se focaram em combater o estado islâmico depois do massacre ao jornal Charlie Hebdo (Jan 7, 2015)
[62] https://storymaps.esri.com/stories/terrorist-attacks/?year=2017
[63] https://en.*wikipedia*.org/w/index.php?title=List_of_terrorist_incidents_in_January_2018&action=edit

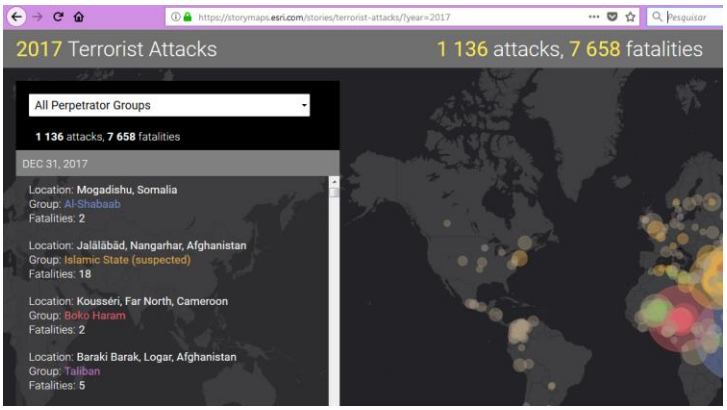


Figura 2- OSINT - ataques terroristas em 2017

2.12 OSINT nos Serviços de Informações e Segurança

As diversas forças nacionais, incluem grupos variados, dedicados a diferentes contextos, e com diferentes modos de actuação. A imagem que se segue contém as forças da Defesa/militares, que operam e geram informações mais geoglobais do que por exemplo, as forças de segurança, que são civis e operam junto das populações e cidadão comum. Por fim, temos os serviços de segurança e informações, também elas consumidoras de informações que podem ser partilhadas, mas serão com certeza diferentes. A polícia judiciária com a sua tarefa de investigação pode precisar ter informações sobre um suspeito estrangeiro que possivelmente o SIS ou o SEF poderão ter, ou mesmo o serviço de informações militar SIED (que não aparece na imagem).
Obviamente, cada força precisa de diferentes informações, que poderão, no entanto, ser partilhadas se assim quiser:

- a vontade política;
- os organismos que geram a informação;
- cada chefe da sua "chafarica" (para não morrer a informação, devido a burocracias)

A falta de partilha de informações em tempo útil e pelas várias entidades, é possivelmente a maior causa para não se poderem fazer detenções com maior rigor e celeridade. É certo que os serviços de informações não terão a mesma capacidade de outros serviços de outros países, devido ainda aos recentes acontecimentos dos serviços de informações no tempo do Estado Novo, mas isso serão águas passadas. Como tudo é preciso controlo e bom senso.

Figura 3-Forças militares, forças e serviços de segurança. Não inclui o SIED/M.

Nota: estão incluídos nos apêndices, conteúdos relativos a estas forças de segurança e o importante papel da OSINT para todas elas.

2.12.1 Órgãos de investigação criminal e forças de segurança

2.12.1.1 O caso da GNR

"A recolha de inteligência nas fontes abertas é uma importante ferramenta para as polícias, embora haja ainda um vazio legal nas definições dos sistemas criminais europeus o que restringe e em certos casos, até leva a abandonar importantes fontes de informação."[64]

Inspector-chefe Rogério Bravo, Polícia Judiciária

Temos nos órgãos de investigação criminal, a Polícia Judiciária, que com um papel mais assertivo e com base em factos concretos, tem a Unidade Nacional Contra o Terrorismo[65] (UNCT) e a mais recente Unidade Nacional de Combate ao Cibercrime e à Criminalidade Tecnológica (UNC3T[66]).

Existem ainda mini-células de pesquisa de informações e OSINT em diversas forças de segurança, pouco referidas diariamente, mas que mostram a crescente importância dada à OSINT.

Senão vejamos por exemplo, **o caso da GNR**:

Em 2015[67], é aprovada a Estratégia Nacional de Combate ao Terrorismo. Nela, é criada ou é dado aval, para a criação da Divisão de Cibersegurança da Guarda Nacional Republicana(GNR). A nova estrutura vai permitir à GNR ter" capacidade *para monitorizar a publicação de conteúdos online de apoio ou incentivo a práticas terroristas, com o objetivo de identificar potenciais ameaças e emitir alertas que darão origem a uma investigação mais aprofundada… A unidade vai fazer a "monitorização e deteção de ameaças, alerta e resposta, em colaboração com as autoridades competentes para a investigação", explicou ao jornal o tenente coronel Paulo Santos, que a vai*

[64] OPEN SOURCES IN CYBERCRIME INVESTIGATION: concept and implications disponível online em http://www.academia.edu/7301278/OPEN_SOURCES_IN_CYBERCRIME_INVESTIGATION_concept_and_implications

[65] https://www.policiajudiciaria.pt/unct/

[66] UNC3T – criada pelo Decreto-Lei n.º 81/2016, https://www.policiajudiciaria.pt/unc3t/

[67] 28 de Abril de 2015, em "Nova unidade da GNR procura apoiantes do terrorismo na Internet, em https://tek.sapo.pt/notícias/computadores/artigos/nova-unidade-da-gnr-procura-apoiantes-do-terrorismo-na-internet

liderar. A criação deste tipo de unidades é uma das principais medidas previstas na Estratégia Nacional de Combate ao Terrorismo e tem em conta o facto de a Internet ser um dos principais palcos para o recrutamento de combatentes por organizações terroristas como o Estado Islâmico… ".

Embora não seja referido acima o uso do OSINT, isto mostra a cada vez maior importância das informações na ajuda à segurança e na prevenção de situações que coloquem em causa o Estado de Direito. Em baixo, ainda a GNR, vai agora dedicar-se especificamente ao OSINT, mostrando com isto, o reconhecimento da sua importância.

Em 2016, segundo o documento do Fundo da Segurança Interna, no sítio web da GNR[68] e Secretaria geral da Administração Interna[69], este órgão aprovou (em 2016), para finalizar em 2018, o gasto de quase 400 mil euros para a aquisição de uma ferramenta de software OSINT. QUATROCENTOS MIL EUROS (sendo que a Europa contribui com 200.000€, e o "contribuinte" 154.000€ totalizando cerca de 754.000€). Objectivos segundo o site:

- *Aumentar a capacidade de recolha e análise de informações, para melhorar a eficácia da prevenção e combate aos fenómenos criminais, terrorismo e criminalidade transfronteiriça;*
- *Assegurar a permanente monitorização, acompanhamento, análise e disseminação de informações públicas, policiais e criminais, em apoio das atividades e operações correntes, auxiliando no processo de tomada de decisão;*
- *Proceder à pesquisa, recolha e tratamento de informações de fontes abertas;*
- *Possibilitar o desenvolvimento de um policiamento orientado pelas informações.*

2.12.1.2 PJ - Polícia Judiciária

A Polícia Judiciária (PJ) é o principal órgão policial de investigação criminal, vocacionado para o combate à grande criminalidade nomeadamente ao crime organizado, terrorismo, tráfico de estupefacientes, corrupção e criminalidade económica e financeira. A PJ está integrada no Ministério da Justiça, atuando sob orientação do Ministério Público.[70]

A PJ tem um novo departamento, a Unidade Nacional de Combate ao Cibercrime e à Criminalidade Tecnológica (UNC3T), que foi criado recentemente e pretende dar resposta a uma nova forma de criminalidade, assente em meios tecnológicos.

Sendo um departamento de base tecnológica e de investigação, é, portanto, um excelente palco de utilização OSINT. Até porque o que consta na web é considerado para todos os efeitos, prova. É também sabido que as informações que "caem" na Internet quando são graves ou atentatórias ao Estado de Direito, têm "tendência a desaparecer" rapidamente, pelo que, se o OSINT não for feito e não houver quem guarde essas evidências no momento, as mesmas poderão desaparecer como se nunca tivessem existido. Essas informações que se refere como caírem na internet podem ser tão graves como a ameaça à integridade humana, ameaças

[68] http://www.gnr.pt/ficheiros/seguranca_interna/3.pdf

[69] https://www.sg.mai.gov.pt/Noticias/Paginas/Fundo-de-Seguran%C3%A7a-Interna-financia-Unidade-OSINT-da-Guarda-Nacional-Republicana.aspx, 16-02-2017

[70]https://pt.*wikipedia*.org/wiki/Pol%C3%ADcia_judici%C3%A1ria

ao Estado de direito, ameaças contra pessoas ou instituições. A internet é também muito usada para *cyber-bullying* entre as gerações mais novas.

Não tendo meios humanos para patrulhar constantemente blogues e *sites* terroristas seja na *web* ou *dark web*, o recurso à OSINT automatizada pode ser um factor decisivo entre apanhar uma informação atempadamente e agir. Ou saber da informação depois de passado o acontecimento e ter de investigar. Ou até mesmo nunca chegar a saber de nada (e não fazer investigação do ocorrido). Quem diz *sites* terroristas diz blogues de crianças/adolescentes, diz sites de encontros, salas de conversação, …

A UNC3T destaca-se para esta tese devido a alguns dos seus objectivos em concreto:

- v) De espionagem, quando cometido na forma de um qualquer programa informático concebido para executar ações nocivas que constituam uma ameaça avançada e permanente.
- 3 – A UNC3T colabora e apoia de forma direta as ações de prevenção, deteção e mitigação desenvolvidas pelas entidades nacionais com competências definidas por lei para a segurança nacional do ciberespaço.
- Elaborar e manter atualizado o Plano Nacional da Polícia Judiciária para a Prevenção e o Combate ao Cibercrime, nomeadamente, em articulação com o Centro Nacional de Cibersegurança;
- c) Assegurar o regular funcionamento de um grupo consultivo informal para debate e aconselhamento estratégico, formativo, jurídico, técnico e científico de questões relacionadas com o cibercrime, com a criminalidade tecnológica e a cibersegurança;
- e) Desenvolver ações de contrainformação criminal;
- f) Dar apoio em ações de carácter técnico para a recolha de prova digital, nomeadamente, ações encobertas e interceção de dados;
- g) Apoiar investigações que exijam conhecimentos técnicos especializados, nomeadamente, redes de anonimização, mercados virtuais, moedas virtuais, análise de programas maliciosos.
- c) Testar e desenvolver ferramentas específicas para a investigação do cibercrime, da criminalidade tecnológica e da decifragem de dados;

2.12.2 Serviços de informações

Portugal tem um organismo da dependência da Presidência do Conselho de Ministros chamado Sistema de Informações da República Portuguesa (SIRP) que tem a seu cargo, dois serviços de informações (chamados normalmente de "as secretas"[71]) e que é composto

- Pelo SIS[72], o serviço de informações que lida com a segurança (civil) e informações internas

[71]Embora chamados normalmente de "as secretas", o simples facto de serem conhecidos, assim como a sua morada, objectivos, etc., torna a definição de secretos, pouco secretos. Segundo o próprio site do SIED, "Os Serviços de Informações trabalham em segredo, mas não são secretos!".

[72]SIS - Serviço de Informações de Segurança

- Pelo SIED[73], o serviço de informações que lida com a Defesa e segurança externas. A este respeito ver o ponto 3.7.1

Ambos serviços estão bastante reduzidos nas suas capacidades de actuação devido ao excesso de poder e abusos, que no tempo da ditadura portuguesa, outras organizações como estas funções, tiveram. Ambos estão na dependência do SIRP, que responde directamente ao Primeiro-Ministro.

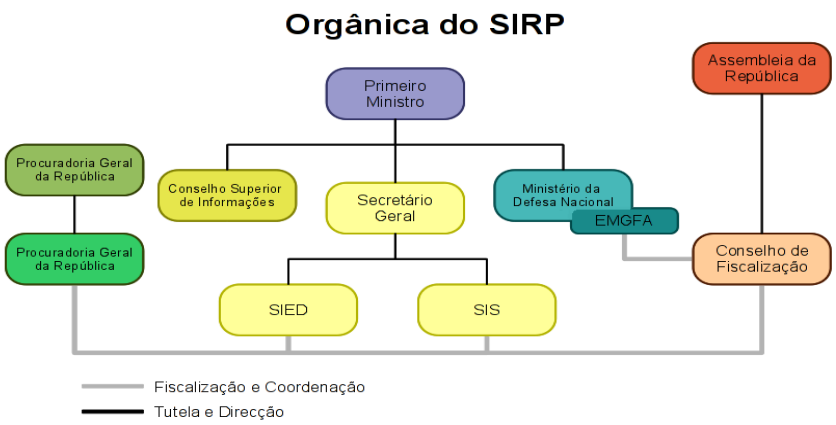


Figura 4- Orgânica do SIRP

Na imagem abaixo vemos a evolução histórica dos serviços de informações portugueses desde o fim da PIDE/DGS até aos dias de hoje (ignorar 2004).

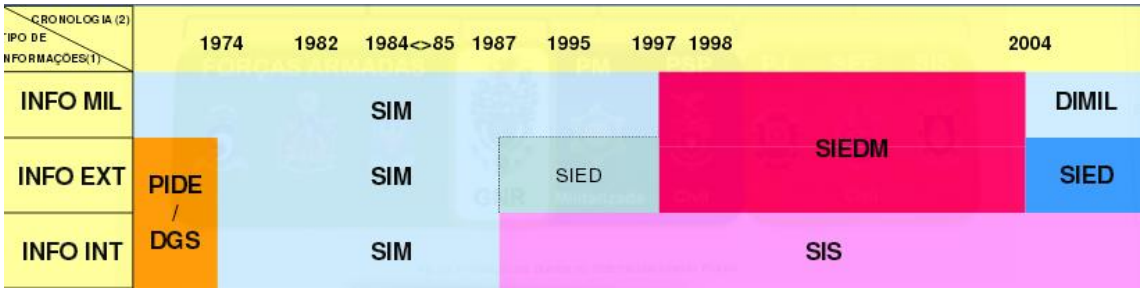


Figura 5- Evolução História do SIRP

2.12.2.1 SIS – Serviço de Informações de Segurança

Os objectivos do SIS, são de acordo com o seu site[74], *ipsis verbis*, os seguintes: "***RECOLHER, PROCESSAR E DIFUNDIR*** *informações no quadro da Segurança Interna, nos domínios da sabotagem, do terrorismo, da espionagem, incluindo a espionagem económica, tecnológica e científica e de todos os demais atos que, pela sua natureza, possam alterar ou destruir o Estado de direito democrático, incluindo os movimentos que promovem a violência (designadamente de inspiração xenófoba ou alegadamente religiosa, política ou desportiva) e fenómenos graves de criminalidade organizada, mormente de carácter transnacional, tais como a proliferação de*

[73]SIED - Serviço de Informações Estratégicas de Defesa
[74]Disponível online em https://www.sis.pt/quem-somos/o-sis

armas de destruição em massa, o branqueamento de capitais, o tráfico de droga, o tráfico de pessoas e o estabelecimento de redes de imigração ilegal."

O SIS refere que pretende combater as ameaças à integridade nacional, abaixo listadas:

- Terrorismo Transnacional;
- Espionagem clássica;
- Espionagem económica;
- Crime Organizado;
- Extremismos ideológicos, religiosos;
- Branqueamento de Capitais;
- Tráfico internacional de Armas de Destruição em Massa (ADM) - Proliferação;
- Tráfico de Seres Humanos e Migrações ilegais;
- Cibercriminalidade;
- Novas Formas de Crime;

Existe a ressalva nesta página web em que o Legislador[75] "*foi aliás muito claro na destrinça dos campos de ação das informações de segurança das da investigação criminal, criando, para as duas áreas, instrumentos distintos: Sistema de Informações/ Sistema de Investigação Criminal e ainda o Sistema de Segurança Interna. É no âmbito do quadro de atuação que em seguida se apresenta (Meios de Atuação, Limitações e Especificidades) que o SIS desenvolve a sua acção…*[76]".

O SIS ainda no mesmo site e na mesma página refere que "*atua em conformidade com os princípios da NECESSIDADE, PROPORCIONALIDADE e ADEQUAÇÃO*", definindo os seus princípios de actuação:

- *Obtendo informações através de fontes humanas, algo que se designa por HUMINT (Human Intelligence);*
- *Acedendo, mediante a celebração de Protocolos com as entidades públicas competentes, a dados e informações constantes de ficheiros dessas mesmas entidades.*

E é aqui que verificamos a importância também do OSINT nos nossos serviços (ainda que não se possua dados para verificar a % deste tipo de dados em comparação com as outras fontes):

- *Processando informações recolhidas através de fontes abertas e documentos não classificados que se encontram ao alcance do público em geral, método designado por OSINT (Open Source Intelligence);*

Os limites[77] deste serviço estão impedidos de:

- *Limitar os direitos liberdades e garantias fundamentais;*
- *Realizar interceções de comunicações;*
- *Deter pessoas;*

[75]Legislador é uma figura abstracta a que se recorre quando nos queremos referir ao criador legislativo das Leis. *Os diplomas emanados da Assembleia da República têm a designação de Leis e os diplomas emanados do Governo têm a designação de Decretos-Lei. https://pt.wikipedia.org/wiki/Lei_de_Portugal*

[76]https://www.sis.pt/quem-somos/o-sis

[77]Lei n.º 9/2007, de 19 de Fevereiro
http://www.pgdlisboa.pt/leis/lei_mostra_articulado.php?nid=910&tabela=leis

- *Instruir inquéritos ou processos penais;*
- *Exercer actos próprios da competência dos tribunais ou das entidades policiais;*
- Operar fora do espaço sujeito aos poderes soberanos do Estado Português; Art.º34

2.12.2.2 SIED/M – Serviço de Informações Estratégicas de Defesa/Militar

O SIED é o serviço de informações português que se assemelha à CIA americana. É o serviço de informações que lida com as ameaças externas. Tem como missão, "produzir informações estratégicas visando a salvaguarda da independência nacional, dos interesses nacionais e da segurança externa do Estado Português"[78]. É sua missão:

- A avaliação da ameaça terrorista
- A identificação de redes internacionais de crime organizado, nomeadamente as envolvidas em narcotráfico, facilitação da imigração ilegal e proliferação nuclear, biológica e química (NBQ);
- O acompanhamento permanente da situação de segurança das comunidades portuguesas residentes no estrangeiro;
- O alerta precoce para situações de potencial comprometimento dos interesses nacionais;
- As matérias políticas, sociais, económicas, energéticas e de defesa que constituam prioridade da política externa portuguesa.

Também o SIED está impedido de ter "competências policiais, estando os seus funcionários, civis ou militares, proibidos de exercer poderes, praticar atos ou desenvolver actividades do âmbito ou competência específica dos tribunais ou das entidades com funções policiais, sendo-lhes expressamente proibido proceder à detenção de qualquer indivíduo ou instruir processos penais." [79]

No trabalho do "Curso de Estado Maior" do Major de infantaria Serra Pedro[80], é referido que o SIS e o SIEDM acumulam funções. Sugere o autor então a "*fusão do SIS com o SIE, ficando este na dependência do MAI".* Sugere em alternativa que devia o SIEDM perder as letras D e M, passando a ser Serviço de Informações Estratégicas, e a sua dependência política passar para o MNE[81].

[78]https://www.sied.pt/quem-somos/o-sied
[79]https://www.sied.pt/quem-somos/o-sied
[80] Sistema de Informações Militares. Contributos para a sua reestruturação e operacionalidade., Maj Inf Para Serra Pedro, disponível online em
https://comum.rcaap.pt/bitstream/10400.26/11894/1/MAJ%20Serra%20Pedro.pdf
[81] MNE – Ministério dos Negócios Estrangeiros

2.13 OSINT – Utilização para ataques informáticos

Já foi referido anteriormente que a OSINT pode ser utilizada no processo de aquisição de alvos e de vulnerabilidades (não só de sistemas, mas também de pessoas), aquando da primeira fase de reconhecimento de um ataque informático. A primeira fase do ataque é a fase de obtenção de informações e é de tal forma importante, que ditam muitas vezes, se vale a pena o ataque ou não. O ataque preparatório com OSINT é baseado no reconhecimento passivo e na descoberta de *assets* ou bens de interesse.

Todas as empresas, entidades, Estados, pessoas comuns, que tenham um site na Internet, estão a oferecer publicamente, quer queiram quer não, informações que podem ser usadas a favor ou contra si. Com recurso a ferramentas, podemos automatizar a recolha de informação desses sites e serviços:

- ataque ao site web: recolha automatizada de emails e nomes de pessoas, análise do código-fonte, localização da empresa, contactos, entre outras fontes que poderão propiciar um ataque posteriormente
- ataque ao utilizador: recolha automatizada de todo o conteúdo do site, podendo este ser copiado ou clonado para atacar os seus utilizadores, enviando emails de spam/phishing/malware e redireccionando as vítimas para um site falso, para, usando ataques diversos, como por exemplo, man-in-the-middle, obter as credenciais de acesso e ataque posterior do site
- ataque ao servidor: enumeração de utilizadores, pastas, ficheiros, serviços em funcionamento. Objectivo é lançar um ataque efectivo e concreto (um ataque geral, para além de demorado, alertaria os potenciais serviços de defesa e bloquear-nos-ia). Interessa-nos aqui saber quais são os serviços, quais os portos abertos, quais as versões, quais as tecnologias utilizadas, entre outros
- ataque ao utilizador: através de engenharia social, é possível fazer-nos passar por um funcionário da empresa do sítio web, e pedir directamente ao utilizador que nos dê algo, ou que faça algo. Pode este dar a password, ou até mesmo facilitar acessos ao sistema. Por exemplo, um atacante que crie um perfil falso de empregador no LinkedIn, pode facilmente obter muitos dados pessoais de uma vítima, fingindo de contas que o quer contratar e pedindo-lhe o CV
- ataque à empresa: através do ataque ao funcionário acima descrito, via LinkedIn, é possível saber que a empresa usa "x" tecnologia, y produtos, e que pode estar a desenvolver uma nova tecnologia. Tudo secreto, mas o funcionário disse estes dados sem se aperceber, para o falso empregador
- *Google maps* e afins: é possível ver como e onde se localiza a empresa-alvo. Onde estão os caixotes de lixo, onde estão as casernas dos vigilantes, como está disposta a empresa, se tem casotas (cães de guarda, portanto), entre outros. Esta visualização permite a posterior tentativa de intrusão, análise de lixo (*dumpster diving*), entre outros. O atacante não precisava do *Google,* mas fazer esta verificação através de imagens aéreas e sem sair de casa, não tem preço.

O OSINT é utilizado na primeira fase dos ataques de pentesting como forma de avaliação do alvo, suas forças e fraquezas. Entre as várias tarefas no ataque preparatório e que podem ter a composição abaixo, o projecto criado para este trabalho, tentou implementar algumas.

Algumas das capacidades do OSINT no *pentesting:*

(*note-se que embora o OSINT não envolva a intrusão nem o contorno de protecções, alguns dos métodos abaixo são activos, mas de forma sustentada de tal modo que o tráfego não causa qualquer restrição nem a pesquisa causa mais tráfego que o que seria normal por um mero visitante, e portanto cai na categoria de ataque passivo*):

Fase de reconhecimento: Investigação de um sítio web para posterior ataque

- Tecnologias utilizadas - O sítio web tem interesse, uma base de dados que pode ter algo… mas para atacar com sucesso precisamos saber o que é que o sítio web utiliza. O que usa pode ter vulnerabilidades que podem ser atacadas. Um servidor web desactualizado? Uma tecnologia antiga que já não tem manutenção e tem falhas de segurança? Uma aplicação instalada de blog esquecida e vulnerável a *SQL injection*? Para sabermos isto temos de fazer uma pesquisa a alguns sítios web que fazem esta pesquisa por nós (estamos a usar OSINT.. não estamos activamente a fazer testes de penetração...). Podemos usar por exemplo, o Wappalyzer[82], ou o builtwit[83]. Ambos fornecem informações muito interessantes. Vejamos um exemplo, o sítio web Ceger.gov.pt utiliza *Windows*, *ASP*, *Jquery* e como servidor Web, o *IIS*. O *builtwith* oferece a mesma informação, inclusive com a versão especifica da tecnologia e uma pequena análise de cada uma, incluindo a 1ª vez que foi vista e a última. *Teremos vulnerabilidades recentes para algumas destas?*

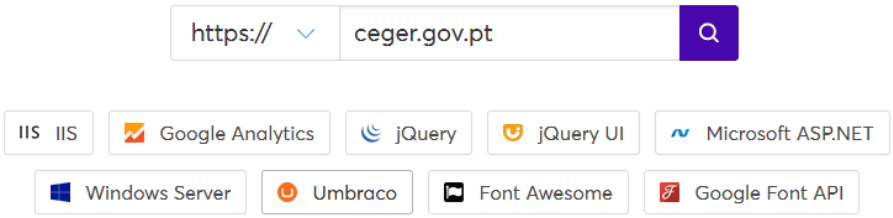


Figura 6 - Obtenção das tecnologias utilizadas por um sítio web

- Alojamento do sítio web – onde está alojado? Quem é o seu "*owner*"? Quem o registou? Qual o IP por trás do DNS? Existem mais sites por trás daquele endereço IP? O que diz o DNS e o WHOIS sobre o endereço?
- Robots.txt – o ficheiro existe? O que tem? Permite *crawlar*? (se não permitir não é ético o *crawl* embora possa ser feito na mesma)
- Clone do site - Para não alarmar o dono do site por estarmos demasiado tempo no mesmo e a pedir páginas consecutivamente, podemos simplesmente obter todo o site. Quando o alarme for gerado, já temos o sítio web todo em nosso poder para análise. O clone do sítio web pode ser usado para enganar outros utilizadores. Por exemplo, com um ataque ARP e MITM.
- Extracção de hiperligações – o sítio web é interessante? Possivelmente as suas hiperligações também serão. Podem levar a outro material de interesse? Se não clonámos o site, então isto deve ser feito.

[82] https://www.wappalyzer.com/
[83] https://builtwith.com

- Actualizações de conteúdo– o sítio web é interessante? Podemos verificar de x em x tempo manualmente ou usar os seus *feeds*, se os houver. Em alternativa podemos usar um *crawler* que nos vá indicando se o sítio web mudou. O interesse reside no conteúdo que pode ser colocado, apagado, alterado… Um sítio web com interesse é por exemplo, um site de vulnerabilidades. Pode acontecer que aquele sistema que queríamos atacar está OK, mas de repente, aparece uma vulnerabilidade…
- Histórico – o sítio web tinha uma informação útil ou um documento que não devia lá estar. Retiraram e já não está visível. É possível o Google ter este conteúdo indexado? Ou, o sítio web está constantemente a ser actualizado.. poderá ter informação antiga relacionada com ex-trabalhadores ou relacionados com assuntos de potencial interesse?
- *Crawl e scrap* – o sítio web alvo é grande, está sempre em mudança ou os conteúdos são dinâmicos e é impossível obter tudo? Recorrendo a um *crawler* bem configurado, talvez seja possível recolher a informação sem baixar todo o site. Especialmente se o sítio web tem produtos e preços. O crawl permite obter apenas o que se deseja.
- Metadados – o sítio web tem documentos? Esses documentos podem conter metadados, como por exemplo o nome do seu criador? Podemos usar essa informação para ataques de engenharia social? Existem imagens? Estas contêm coordenadas GPS? Podemos obter passes de ficheiros ou dicas para tal?
- Certificados – o sítio web tem certificados digitais? Estão válidos? Quem o emitiu? São certificados que usam cifra baixa e que podem ser atacados para ataques MITM?

A maior parte das ferramentas que permitem fazer estes ataques são open source, ou de código-aberto, partilhado por hackers e pessoas "comuns" que desejam partilhar conhecimento. Não confundir *open source* com *open source* de OSINT.

Através de OSINT, e da pesquisa sistemática de *posts* no *PasteBin*, *GitHub*, entre outros, é possível associar *hackers* maliciosos, a *malwares*. Os pequenos erros de programação ou de envio de emails (*spam*/*phishing*) permitem muitas vezes a visualização de nomes, comentários, às vezes até o email, do atacante que desenvolveu e disseminou o *malware*.

A solução não é universal, nem 100% eficaz, mas a educação de todos os trabalhadores e colaboradores, para o perigo, permite reduzir a área de ataque.

2.14 *Intelligence* - A necessidade de uma melhor OSINT

Robert David Steele[84] é considerado um dos grandes impulsionadores da OSINT. Em entrevista[85] para o site FrontLine[86] afirma que embora tenha sido o criador e o responsável pela criação e contratação de pessoal de *intelligence* para a Marinha, descobriu que o que mais precisava para produzir *intelligence* não era secreto e não estava disponível no Governo (CIA): estava no privado. Um dia, fez-se um exercício: R. Steele contra a inteligência internacional americana (Comissão Aspin-Brown) sobre o tópico do Burundi.

[84] *Ex-oficial de inteligência e infantaria, ex-espião, fez parte da CIA, como chefe.*

[85] www.pbs.org/wgbh/pages/frontline/shows/hackers/interviews/steele.html

[86] http://www.pbs.org/wgbh/pages/frontline/

Steele apenas com seis telefonemas, conhecendo o meio, conseguiu obter mais informações que a Comissão. Os telefonemas foram para seis elementos diferenciados e que obtiveram diferentes meios (cartas topográficas, movimentações militares, cultura local, …) sobre o assunto em causa (Burundi). "Ganhou" obtendo mais informações. Não utilizou segredos nem espionagem.
Segundo Steele só nos anos 90, a Internet realmente cresceu e passou a ser uma boa fonte de divulgação de informação, com comunidades específicas e melhoria de procura de informação. Passou a haver um melhor balanceamento entre as pessoas e os Governos.

No sítio Web[87] da *Forbes* em 2006, Steele critica os serviços de informações:

- Falharam na protecção dos EU relativamente aos aviões do ataque de 11 de Setembro de 2001
- Bastaram uns telemóveis e gente consciente para se salvar um avião: "*the only hijacked airplane that failed to hit its target on Sept. 11 was the one where informed citizens were able to take direct action. It gave proof that our national security establishment is broken. A $500 billion per year defense department and a $50 billion per year secret intelligence community failed where a few brave citizens armed only with cell phones succeeded*.
- Maior confiança no "olho do satélite" do que no olho humano
- As informações devem ser utilizadas para tomar boas decisões a todos os níveis e não apenas para roubar segredos industriais

Em resposta aos acontecimentos acima, Steele:

- Pede a reinvenção dos serviços de inteligência
- Explica que para uma nação sobreviver às tribulações futuras, cada cidadão deve ser simultaneamente um colector e consumidor de inteligência, apto a divulgar adequada e em tempo real.

Aponta uma ideia central: a inteligência colectiva

1. O poder do pensamento combinado e colectivo de grupos de pessoas.
2. Quando se escolhe um candidato pelo partido, quando pessoas estranhas se juntam para resolver problemas comuns, quando se editam entradas em enciclopédias pela internet.
3. O mesmo se vê nas formigas que são capazes de manter "ninhos" complicados e executar ataques militares enormes, muito além das capacidades intelectuais de qualquer uma.
4. Quando em 2004, após o tsunami as pessoas enviaram vídeos e fotos, e se entreajudaram para recolher e identificar e ajudar pessoas.

Aponta para um caminho: o da criação de uma agência de *intelligence "Open Source"*

- *How can we use this to reform intelligence? I suggest we create a national Open Source Agency. Half of the money earmarked for the agency would go toward traditional intelligence work. The other half would provide for 50 state-wide Citizen Intelligence*

[87] http://www.forbes.com/2006/04/15/open-source-intelligence_cx_rs_06slate_0418steele.html - 4/19/2006 @ 9:00AM

Networks, including a 24/7 watch center, where citizens can both obtain and input information.

A verdade é que a necessidade de OSINT foi repensada, e foi efectivamente criado, como já referido, o Directorado OSINT.

Como vimos no capítulo 10, em 2005, a CIA irá criar este Centro de Open Source.

2.15 O caso de Tancos. OSINT e *fake news*

Chegou pela comunicação social (jornal online El Español[88]), a informação que teriam sido roubadas (furtadas é o termo mais correcto), da base militar de Tancos, armas e munições, entre outros[89]

O Ministro da Defesa da altura, Azeredo Lopes, quando questionado, disse que teve conhecimento do "potencial" furto em Tancos através da comunicação social. O mesmo ministro também referiu que até admitia nem ter havido furto nenhum[90], quando o jornal espanhol até enumerava o que tinha sido furtado[91].

Temos aqui um exemplo do OSINT (*media* tradicionais) a funcionar mais depressa do que os serviços de informações e outros meios "oficiais" e "oficiosos", ao próprio ministro.

Potencial "*fake news*"?: "*Sabemos que não há nenhum relatório que tenha sido produzido pelos serviços de informações, quaisquer que eles sejam. E é importante saber quem foi, com que motivação foi fabricado esse documento, falsamente atribuído aos serviços de informações", afirmou à agência Lusa Azeredo Lopes*."[92]

"*O Exército Português não quis divulgar a lista de material de guerra roubado em Tancos mas o jornal El Español publicou este domingo toda a informação, incluindo quantidades"*[93]. Aparentemente o jornal sabia mais do que o próprio exército… E veio parar à internet mais depressa que ao Ministro da Defesa.

Para a ferramenta proposta neste trabalho, iKNOW, a detecção desta notícia, sua análise e processamento, por terem sido detectadas as palavras-chave "sis, sied, polícia judiciária, terrorismo", dar-nos-á através do cálculo das métricas presentes, um "bom" resultado que fará

[88] https://www.elespanol.com/

[89] Lista de armas, munições e outros roubados no site do Jornal de Negócios, disponível online em https://www.jornaldenegocios.pt/economia/defesa/detalhe/jornal-espanhol-divulga-lista-de-material-roubado-em-tancos

[90] "No limite, pode não ter havido furto nenhum. Como não temos prova visual nem testemunhal, nem confissão, por absurdo podemos admitir que o material já não existisse e que tivesse sido anunciado... e isto não pode acontecer", Ministro da Defesa Azeredo Lopes em 10-9-2017

[91] https://www.jornaldenegocios.pt/economia/defesa/detalhe/jornal-espanhol-divulga-lista-de-material-roubado-em-tancos

[92]Jornal público online, https://www.publico.pt/2017/09/26/politica/noticia/ministerio-da-defesa-pede-divulgacao-na-integra-de-alegado-relatorio-de-secreta-militar-1786825, consultado em 12/12/2017

[93]http://observador.pt/2017/07/02/espanha-divulga-a-lista-de-material-roubado-em-tancos/, consultado em 12/12/2017

armazenar esta página *web*, seguir hiperligações circundantes e introduzir o assunto principal no relatório final.

Nota: o tema das *fake news* é um tema sério que pode inclusivamente fomentar guerras[94]. O impacto das fake news é grande e deve ser combatido. Devido a restrições de espaço, encontra-se nos apêndices, o capítulo "2. *Fake news*: propagação, impacto e combate" dedicado a este tema.

2.16 Redes sociais – "Adivinhar" perigos e evitá-los

Uma das grandes invenções dos últimos anos, foram as redes sociais. O sucesso foi tão grande que se popularizou e hoje tem milhões de utilizadores no mundo todo. A China é um dos países que censura estas redes sociais, usando apenas uma sua, muito limitada e fechada ao resto do mundo.

O Facebook em Portugal é a rede social com mais utilizadores, que podem criar as suas próprias páginas (públicas ou reservadas) e permitir a outros introduzir notícias, comentários e gostos/*likes*. Páginas essas que têm como tema os mais diversos assuntos, e que não têm uma supervisão, estando por isso à liberdade de cada um colocar basicamente, o que bem lhe entender, grande vigilância, sem problemas sociais, sem punições aparentes.

Em poucos anos, esta rede social começou a ser uma zona onde se fala de tudo, inclusive, serve de plataforma de discussão de greves, manifestações, divulgação de crimes, divulgação de informações pessoais e de terceiros, fotos variadas, invasões de privacidade, *cyberbullying*[95], palco de discursos de ódio, seitas, e até como preparação de ataques a pessoas e/ou a sistemas informáticos.

Ataques como os que ocorreram nas datas de 25 de dois anos seguidos, foram anunciados previamente na rede social *Facebook*. Também por experiência própria, se verificou que após notícias em que havia algum tipo de acção das polícias mal vista no *Facebook*, na própria tarde e noite, ocorriam ataques e/ou actos preparatórios(*portscans*) para tal.

[94] "How Fake News Could Lead to Real War", disponível online em https://www.politico.com/magazine/story/2019/07/05/fake-news-real-war-227272?fbclid=IwAR25M5lBPv0_oJYTRbK-DDZJHM9pTEBt1HS7M-MX9Ka24sl4ktfDLRGBEJk

[95] *Cyberbullying* – perturbar online fortemente as pessoas através de ameaças físicas, ridicularizações recorrendo à raça, sexo, religião, cultura. Humilhações ou até mesmo roubando a identidade, criando perfis falsos e fazendo-se passar pela vítima, em conversas online com terceiros. Mais informação e sugestões de prevenção em https://www.internetsegura.pt/riscos-e-prevencoes/cyberbullying

Figura 7- Preparação de manifestação anunciada nas redes sociais

O *Facebook* permite também, caso haja indícios que o permitam, fazer pequenas investigações pelas pessoas que estão habilitadas para o fazer, desde a procura de indícios de riqueza externa, álibis, locais frequentados, pessoas conhecidas em fotos, pessoas "amigas" na rede de contactos, entre outras coisas.

A interligação de redes sociais, com outras plataformas, permite, por pessoas mal-intencionadas, ataques preparatórios contra outras. Descobrir desta vez fora-da-lei, se o alvo está a passar férias fora, onde vai almoçar/jantar/sair, se tem filhos, qual o carro, gostos pessoais, contactos, entre tantos outros. Para efeitos de engenharia social, ciber-extorsão, furtos a residências, o *Facebook* e outras redes sociais, vieram trazer uma riqueza de novos vectores de ataque, e o melhor de tudo, é que é tudo OSINT. Informação útil e de valor sem recurso métodos invasivos. Não cabe neste trabalho, dizer em pormenor o que fazer e como evitar, mas é simples, quanto menos informação estiver exposta publicamente, melhor.

2.17 Pesquisas em motores de busca (genéricos a especializados)

Existem diferentes motores de busca cujo objectivo é genérico, como por exemplo o Sapo, o super-conhecido e venerável Google, entre outros. Embora genérico, o Google permite utilizar alguns "truques" (conhecidos por *Dorks*) para pesquisas mais profundas e que podem ser por exemplo, documentos que não deviam estar acessíveis ao público, listas de pessoas, passes, emails, bases de dados ou seus backups, entre tantas outras coisas, que o Google vai reunindo(indexando) permanentemente. Existem dezenas de *dorks* diferentes, e que permitem a um atacante sem grandes conhecimentos informáticos, reunir uma listagem grande de sites que estejam vulneráveis por exemplo a ataques de *SQL injection*[96] [97]. A procura de ficheiros PDF é também de grande utilidade (ex: escrever "openbsd filetype:pdf" no Google, retorna livros, papers e outros documentos em formato pdf sobre o sistema operativo OpenBSD[98]). Os Google *dorks* são poderosos, sabendo o que procurar…. É possível até localizar ficheiros que já foram apagados mas continuam na *cache* do Google. Também é possível por exemplo procurar por frases que tipicamente existem nos sistemas alvo, como por exemplo, câmaras web: *inurl:top.htm inurl:currenttime*

[96] "Latest Google Dorks List 2018 For Ethical Hacking and Penetration Testing", 2019, 2 de Março de 2019, https://gbhackers.com/latest-google-dorks-list/

[97] "*New Google Dorks List Collection for SQL Injection* –SQL Dorks 2019", 2 de Março de 2019, https://gbhackers.com/latest-google-sql-dorks/

[98] http://www.openbsd.org

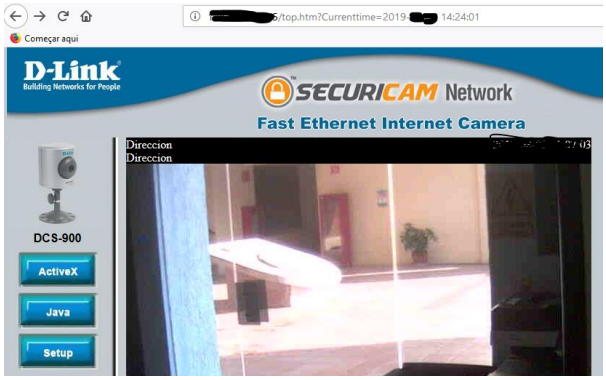


Figura 8 - Google Dorks: descoberta de câmaras web desprotegidas

Nos dias de hoje, quando usamos alguns *dorks* começamos a ser brindados pela Google, pelo que o uso de VPN para usar alguns *dorks* é recomendado (em baixo utilizou-se: "*not for distribution" confidential*), se não queremos ficar na lista de endereços de ips suspeitos, da Google:

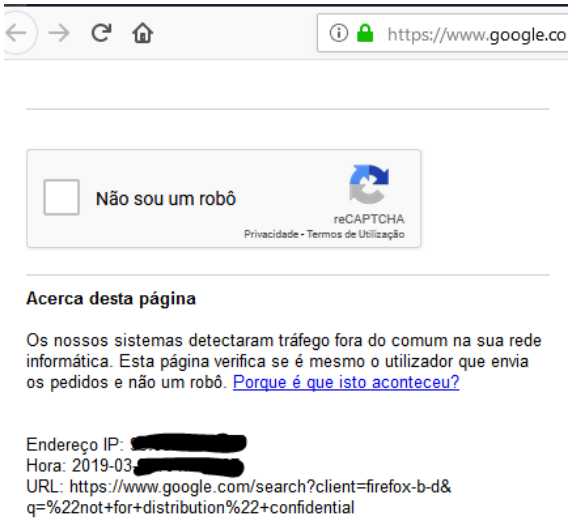


Figura 9 - Aviso do Google relativamente ao uso de Dorks

Não pretendendo ser um motor de busca concorrente aos existentes, temos sites como por exemplo, o *Instya*[99], que nos permitem pesquisar online simultaneamente em vários motores de busca, abrindo uma "janela" para cada. A utilidade e referência a esta ferramenta, está na pré-selecção, integração e bom funcionamento dos diversos motores de busca. Facilita.

Existem também motores de busca especializados que nos permitem fazer buscas mais específicas, como por exemplo, motores de busca académicos como o "Google Académico"[100], o *Social Science Research Network*[101] ou o *Science Research*[102]*(sítio web fantástico e que pessoalmente recomendo devido aos papers e a uma quantidade de informação incrivel, que se conseguem encontrar).*

As próprias redes sociais podem ser utilizadas como motores de buscas, orientadas não a assuntos, mas a pessoas. Mas os grupos, como por exemplo do Facebook são uma fonte enorme de informação, já que estes são "especializados" em assuntos especificos. Estes grupos, foram por diversas vezes fonte de informação relativamente a ataques *hackers* e a

[99] http://www.instya.com/#/web
[100] https://scholar.google.pt/
[101] http://www.ssrn.com/en/
[102] http://www.scienceresearch.com/

manifestações públicas, indicando datas, forma, autores morais, e localização do ataque/manifestação.

Quando se pretende pesquisar por "dados dentro dos dados "(como por exemplo, fotografias), recorremos a motores de busca de Meta-dados. Encontram-se por vezes demasiados dados, mas nem sempre, os que se procuram. Para ver metadados de uma fotografia por exemplo, existem já alguns sites[103] que o fazem. No âmbito deste projecto, foi feito no iKNOW, uma página web, que nos permite enviar uma fotografia, e esta depois devolve-nos as coordenadas GPS, integradas num mapa, (poderia ser disponibilizada toda a informação da imagem, caso tivesse sido esse o objectivo).

Para ataques preparatórios são normalmente usados serviços pagos que, ou já fizeram crawls antes, ou que têm informações sobre vulnerabilidades antigas ou actuais. Para estes casos, existem diversos, destacando-se o SHODAN[104] como um serviço e motor de busca especial, que nos permite fazer uma pesquisa por computadores específicos, *routers*, servidores, portas, entre outros usando filtros. É também chamado de motor de busca de *banner* já que pesquisa também pelo que os servidores respondem aos clientes: dados, opções, informações variadas, serviços que oferecem. As pesquisas podem ser feitas por país, *hostname*, sistema operativo, porto, versão, entre outros. O exemplo mostra a nossa procura por um serviço bem útil: SSH (serviço de acesso e gestão remotos). Poderíamos também procurar por servidores com falhas ou por *webcams* abertas ao público (com ou sem conhecimento dos seus donos…).

Também para ataques (desta vez não preparatórios), existem sites especializados, totalmente OSINT, de onde podemos retirar informações para ataque/defesa de equipamentos, como por exemplo, o Exploit-Database[105], uma base de dados gigante com milhares de vulnerabilidades para diferentes equipamentos, plataformas web, sistemas operativos, …, e actualizada diariamente pela comunidade.

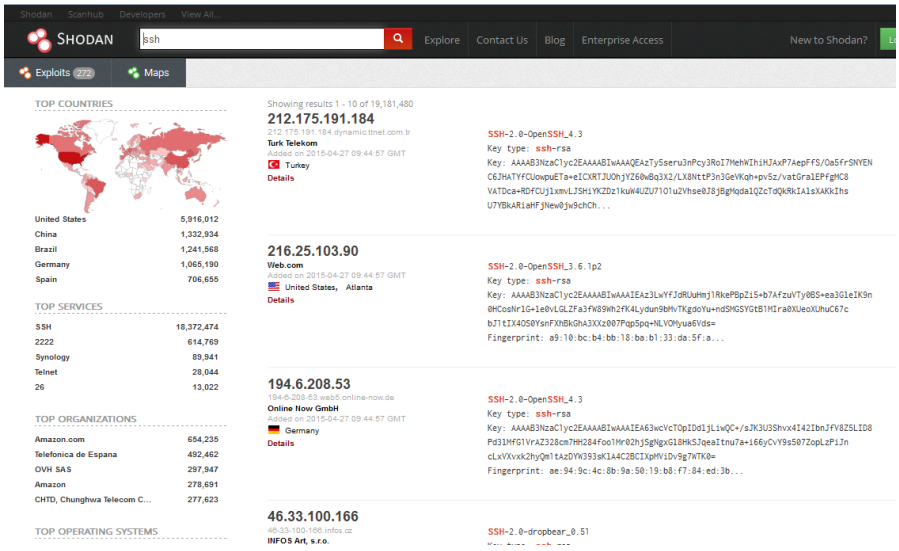


Figura 10 - Motores de busca especializados: procura de equipamentos com vulnerabilidades utilizando o SHODAN

Se o que procuramos são *leaks*, ainda que possivelmente não de forma legal, nem com valor legal, ainda que possam ter sido obtidas legalmente via OSINT. Existem[106] alguns sítios web que

[103] Entre eles, destacam-se http://regex.info/exif.cgi, http://camerasummary.com/, http://www.exifviewer.org/

[104] https://www.shodan.io

[105] https://www.exploit-db.com/

[106] https://data.occrp.org, https://panamadb.org, https://wikileaks.org, https://www.databases.today, https://weleakinfo.com, https://cryptome.org

foram surgindo muito recentemente e cujo título geralmente está relacionado com o conteúdo do leak que lhes deu origem, como por exemplo, PanamaDB, OffShoreleaks, Wikileaks, OCCRP Data, entre outras. Alguns sites de leaks, saem da Internet (acção das autoridades) em pouco tempo. Finalmente, um site que nos permite colocar o nosso endereço de email (ou das vitimas/pessoas que queremos defender), é o *have i been pwned*[107], devolvendo se o nosso email consta/constou ou não de sítios web que foram "*leakados*".

Por fim, *a Deep Web,* que como se sabe, não permite a utilização sem medidas extra, de navegadores web comuns. Nesta "rede" a informação não é ainda anexada, pelo que não temos muita informação de como chegar às informações e serviços pretendidos. Existem alguns motores de busca, com alguma informação, mas ainda não existe algo semelhante a um Google.

2.18 *Pastebin* e afins – Partilha de informações, anonimamente

Uma das ferramentas mais conhecidas para a transmissão para a comunidade e sem responsabilização, é o *PasteBin*[108]. O site fornece um local onde qualquer um pode colocar texto e ligações para outro site, de forma anónima. Mas o texto colocado também pode ser código-fonte de *software* ou informações obtidas de empresas ou pessoas, de forma lícita (ou ilícita), os chamados *leaks*. A pesquisa de informação é muito simples. Basta saber o que procurar ou colocar directamente o endereço que obtivemos sozinhos ou por terceiros. Neste caso o endereço foi o http://pastebin.com/RniQXzqx, mas existem milhares de outros e de outros assuntos. Caso prático: pretendeu-se saber mais informação sobre uma operação detectada nas redes sociais de nome operação "Charlie Hebdo". Colocou-se Charlie Hebdo no *pastebin* e obteve-se de imediato alguns resultados. O que nos interessa: "#charlie hebdo". Obteve-se:

1) O nome da operação #Charlie Hebdo

2) A identificação do grupo e dos seus objectivos nesta operação. Podem estar envolvidos mais grupos. Há também uma forma de contacto e de associação. Geralmente de simples acesso no 1º passo. Os outros já exigem mais medidas e mais técnicas.

> IF YOU HAVE ANY QUESTIONS, PLEASE JOIN US ON https://webchat.anonops.com/
> IRC: irc.anonops.com Port: 6667 Port SSL: 6697 #francophone & #OpCharlieHebdo // https://webchat.anonops.com/
> Twitter: @OpCharlieHebdo // #JeSuisCharlie

3) Um tutorial passo-a-passo dos métodos que podem ser seguidos para colaborar com o grupo. (neste *post* em específico, foram inseridos o código completo e as instruções de utilização. Embora algo complexo, as pessoas só têm de copiar e colar).

[107] have i been pwned, disponível online em https://haveibeenpwned.com/
[108] Disponível *online* em https://pastebin.com/

4) Uma lista de alvos para ataques *DDOS*[109] (e se possível, *deface*[110])

HITLIST ddos // deface :	
WEBSITE⤵	DNS⤵
http://www.joinalqaeda.com/	192.254.235.178
https://shamikh1.info/vb/	109.163.232.173
http://dwl-is.appspot.com/	109.163.232.173
http://online-jihad.com/	192.0.81.250
http://opcharliehebdo.com/	198.100.149.36
http://islam2012.be	23.236.62.147
http://islamic-state.ga/	5.231.64.77
http://ikhwanonline.com/	69.172.201.19
http://ansar-alhaqq.net/	69.172.201.19
https://isdarat-tube.com	104.28.10.110
http://shahamat-arabic.com/	104.28.17.10
http://shahamat-urdu.com	104.28.14.113
http://shahamat-farsi.com	104.28.11.102
…	

5) Imagem da informação e página obtidas.

Figura 11 - PasteBin - indicação dos alvos e instruções para ataque

A tabela abaixo mostra a lista de instruções para o ataque recolhidos no site acima. Veja-se o grau de pormenor colocado:

- Nome da operação e Objectivos ("*shut down Jihadist hate-spreading websites*")

[109] *DDOS – Distributed Denial of Service* – ataque informático distribuído por centenas de sistemas, que pretende fazer a "negação de serviço": o objectivo é tirar do ar (colocar *offline*), sítios web e/ou serviços.
[110] *Deface* – ataques informáticos com um objectivo geralmente político, disseminando mensagens contrárias ou contra o site. Também pode ser usado o ataque para mostrar a perícia do atacante (troféu).

- ONDE, COMO e com QUEM falar (através do IRC, com uso de SSL (comunicação servidor-cliente cifrada, para não ser possível através de técnicas de *sniffing*[111], visualizar o que se está a dizer)
- Como e onde: instruções para: baixar as ferramentas de ataque (sistema operativo Linux específico), consultar sites de instrução da ferramenta, escolher o alvo e verificar se o ataque depois de executado surtiu efeito prático.

"WHAT CAN I DO?"
This is a step-by step guide how to execute a basic DOS attack to shut down Jihadist hate-spreading websites.
Please, educate yourself on the matter first. Ask questions. Share.
NOT everyone should be involved in #OpCharlieHebdo.
If u choose to be, KNOW the risks entitled.

IRC: irc.anonops.com Port: 6667 Port SSL: 6697 #francophone & #OpCharlieHebdo // https://webchat.anonops.com/
Twitter: @OpCharlieHebdo // #JeSuisCharlie

1) Download and install Virtualbox. This program allows you to run multiple operating systems on your computer.
https://www.virtualbox.org/wiki/Downloads
2) Download and install Kali Linux. This is "The most advanced penetration testing distribution".
https://www.kali.org/downloads/
3) Inside Kali, browse to and copy the Slowloris script into slowloris.pl file
http://ha.ckers.org/slowloris/slowloris.pl
4) Browse the hitlist, pick a target.
http://pastebin.com/BB6hzy8D
5) Test to determine timeout.
./slowloris.pl -dns www.example.com -port 80 -test
6) Execute.
./slowloris.pl -dns www.example.com -port 80 -timeout 2000

Note: If you know that the server has a timeout of 3000 seconds, but the the connection is fairly latent you may want to make the timeout window 2000 seconds and increase the TCP timeout to 5 seconds. The following example uses 500 sockets. Most average Apache servers, for instance, tend to fall down between 400-600 sockets with a default configuration.The smaller the timeout the faster your attack will succeed (It will consume all the available resources as other sockets that are in use become available) The closer you can get to the exact number of sockets, the better, because that will reduce the amount of tries (and associated bandwidth) that Slowloris will make to be successful.

./slowloris.pl -dns www.example.com -port 80 -timeout 2000 -num 500 -tcpto 5

Slowloris has no way to identify if it's successful or not though. Use https://isitup.org/ to check if a server is up.

O exemplo acima é apenas um de muitos outros sites do género, destacando sem qualquer preferência, o *HasteBin*[112], o *Pasted*[113], o *CryptBin*[114]*, o PasteLink*[115]*. O ZeroBin*[116] é útil devido à possibilidade de se auto-apagar a mensagem, decorrido x tempo (definido por nós).

2.19 GitHub – Obtenção e partilha de código-fonte (software)

[111] *Sniffing* – neste caso específico, a ferramenta captura dados de tráfego de rede, e consegue verificar o seu conteúdo. Se estiver cifrado, os dados obtidos não são facilmente visíveis, exigindo um decifrar (o que em termos de tempo e esforço, podem não valer a pena para quem tenta).
[112] https://hastebin.com/ - Muito simples de usar. Anónimo.
[113] http://pasted.co/ - Vantagem: a partir de certo valor o site paga a quem colocou e deu visitas. Desvantagem (imensa): privacidade.
[114] https://cryptbin.com/ - simples e cifrado
[115] https://pastelink.net/
[116] http://sebsauvage.net/paste/

O que tem código-fonte e o *GitHub* a ver com *Open Source Intelligence* se o *Open Source* de *Software* são coisas tão distintas? Bem... para *software*, o *GitHub*[117] é uma ferramenta incrível. Permite a qualquer pessoa, sem gastar dinheiro, criar e participar em comunidade, em projectos de *software*. O único senão (e é um senão pequeno) é não ter qualquer controlo sobre o destino do código-fonte criado (estes projectos são de uso público, a menos que se pague anualmente). No site do *GitHub* é possível encontrar projectos de tudo, inclusive:

- ferramentas que já pertenceram a serviços de informações
- *malwares* que foram disponibilizados ao público, a fim de serem estudados (e quiçá melhorados) pelo público
- projectos de Hackers
- projectos científicos de comunidades, estudantes
- …

No âmbito desta tese e do tema OSINT, os temas mais interessantes vão sem qualquer dúvida, para tudo o que permita a obtenção de informação. Através de uma utilização "diferente" e em que "exploramos" o próprio *GitHub*, podemos:

- obter nomes de pessoas (e saber que projectos e interesses tem)
- obter passes (as pessoas esquecem-se que o que usam em casa, é depois enviado para o GitHub, indo as passes no meio do código). Passes comuns, utilizadas para fins pessoais e para fins de serviço, projectos, etc.
- obter *social network intelligence*. Quem contribui para que projecto? Qual o interesse de x numa ferramenta antes utilizada por serviços de informação? Porque razão x membros estão a duplicar projectos de *malwar e (spyware, plat*aforma de criação de *phishing,* vírus, cavalos de Tróia…)? Estão a criar alternativas ou a melhorar/piorar o *malware* existente?

Tudo isto são informações importantes no âmbito do combate às ciber-ameaças, e à utilização da informação para os nossos objectivos. Como curiosidade, parte do iKNOW está no *GitHub* (leitor de *feeds* e *script* para automação da recolha de assuntos de jornais).

Um projecto/repositório de interesse no GitHub e que mostra as motivações de alguns dos seus utilizadores: *GitMiner*[118] - *"Demonstrates the fragility of trust in public repositories to store codes with sensitive information."* Uma das suas utilizações é a pesquisa por ficheiros de configuração dos blogues *Wordpress*, que contenham palavras-passe no seu interior. O site tem o seguinte filme[119] que mostra como usar o *gitminer* para descobrir *hosts*, nome de utilizador e passes, para invadir sistemas via SSH.

O acto de invadir um sistema é ilícito mesmo que a forma como foram descobertas as passes tenha sido feito de forma legal.

[117]https://github.com/

[118]Disponível online no GitHub em https://github.com/xSploited/GitMiner

[119]Video de como invadir um sistema via conexão remota SSH:
https://www.youtube.com/watch?v=yIJOlKZwQQw

2.20 Ameaças à privacidade e anonimidade

Somos confrontados todos os dias, com notícias de quebras de privacidade, roubos de informações de Governos[120], *leaks* de informações, empresas reputadas[121] [122] e clubes de futebol[123] que todos julgavam bem protegidas. Recentemente, o surgimento de novas leis de protecção de dados, GDPR, foi uma boa notícia e uma boa tentativa para obrigar a proteger a privacidade.

E se a maior parte destes acontecimentos acontece lá foram nacionalmente também os há. O sítio Tek, dá conta que a polícia judiciária vai adquirir equipamentos que permitem o acesso de dados de telemóveis à distância[124]. A PJ é uma polícia de investigação que pode precisar desses recursos, e podemos duvidar de quem usar estes equipamentos em quem, mas, temos de acreditar que o fazem para apanhar criminosos.

Existem diversas formas de se localizar um indivíduo na internet, umas mais fáceis, outras mais complicadas. Tentaremos no âmbito do OSINT, mostrar algumas, tentando também mostrar que o que se investiga, pode ser usado também para nossa protecção.

2.21 Como sabem eles? Factores de reconhecimento e identificação online

A inteligência OSINT é feita quase na sua totalidade de forma passiva, descobrindo o que existe e alguém colocou ao alcance de todos, mesmo que possa sido sem querer, mas sabemos que quando algo cai na Internet, ou quando algo é publicado, ou passado na televisão, mesmo que haja esforços para se retirar ou corrigir, o mal já está feito e alguém há-de ter ou lembrar-se e há-de haver uma prova disso registada algures.

A utilização de VPNs ajuda muito na protecção da privacidade, quer no café onde usamos wifi (não usamos, mas é para exemplo), quer junto do ISP, quer onde quer que seja. No entanto, a VPN não nos protege totalmente se formos pouco cuidadosos a lidar com a nossa informação pessoal e com a interacção com os outros utilizadores online.

A descoberta da identidade online pode ser obtida de diversas formas. Esta informação é útil para nos proteger, mas também para tentar identificar alguém e pode ser feito em qualquer sítio, pois é feita com dados existentes.

2.21.1 Identificação de utilizadores

[120] Governo alemão, com roubo de dados na ordem dos 16 gigas.

[121] *Panamá papers* - um escândalo que envolveu muitas empresas e paraísos fiscais, fugas ao fisco...

[122] Samsung, Sony, falha do Googl

[123] *Football leaks*, divulgação de emails e informações confidenciais de diversos clubes de futebol portugueses

[124] Disponível online em https://tek.sapo.pt/notícias/computadores/artigos/pj-vai-conseguir-ter-acesso-aos-dados-de-telemoveis-a-distancia, acedido em 20/02/2018

Os utilizadores podem ser identificados pelo que são, pelo que fazem, pelo que aparentam ser, pela forma como são vistos, pelo sítio onde estão, mas também pelo contrário de tudo isto. Tudo é característico e serve para identificar, mesmo o que não aparentemente é identificável. Algumas formas de identificação de utilizadores:

- Presença prolongada na rede e em sites/serviços
- Visibilidade alta, devido a actividade (posts, tráfego por tempo ou quantidade…)
- *Background* em redes sociais, motores de busca, sites de emprego, etc., quer seja por colocarmos o CV online, quer seja por termos um *blog*, pertencermos a um partido político, equipa de futebol, clube de xadrez ou por alguém nos reconhecer numa rede social ou indicar o nosso nome.
- *Cookies*[125], que podem dar ao dono do site, informações tão diversas e específicas do visitante e seu navegador web que permitem descobrir que um visitante da TOR, também visitou o mesmo site, alguns minutos atrás.
- Análise de tráfego de rede, feito através de *sniffers* de rede, na própria casa ou organização, ou feito de forma profissional por *SOCs*, *CSIRTs* e *hackers*
- Redes abertas, seja ela um café, uma casa com rede aberta, um ponto de acesso aberto propositadamente para captura de dados, …
- Forenses – captura de equipamentos e comunicações

Uma análise às comunicações acima, permite a leitura do analista/hacker/policia, de todas as comunicações não cifradas (que não usem certificados nos sítios web (SSL/TLS) ou que não usem protocolos de comunicações seguras (SSH por exemplo)). As comunicações não-cifradas permitem obter passes, nomes e *cookies* (dados de autenticação, id da sessão, dados do navegador, entre outros).

2.21.2 Identificação de utilizadores – via malware

As actividades expostas na Internet podem levar a que as redes internas possam ser comprometidas a partir do site exposto, via uso de vulnerabilidades como por exemplo falhas de programação, exploradas através de *exploits*. A organização/indivíduo também pode ficar vulnerável se visitar um site malicioso, ou se fizer downloads de *malware* (não propositadamente, mas que venha junto com um software fidedigno por exemplo), ou através da infecção porque funcionário x recebeu um email(*phishing*) com um anexo, e o abriu. O anexo continha um executável que ao ser executado fez a identificação da máquina, utilizador, dados, e os enviou para um servidor malicioso.
Esta táctica de intrusão e detecção não é de todo OSINT e constitui diversos crimes.

2.21.3 Metadados, e o caso das secretas portuguesas

[125] Cookies – pequenos ficheiros electrónicos, enviados pelos sítios web que visitamos, para o nosso computador através do navegador/*browser*. Pode guardar preferências, dados pessoais, dados da nossa visita como por exemplo a data, e volta a ir buscar estes dados na próxima visita. Também o dono do site regista o nosso endereço IP, id do cookie, entre outras informações de potencial interesse.

Os metadados normalmente não mereceriam grande destaque, mas devido a ser um tema recente à escala nacional, tem aqui um mini-capítulo, por se julgar de interesse. Os metadados como anteriormente se referiu, não são os dados em si, mas as informações que acompanham, por exemplo, uma fotografia pode ter metadados como seja a máquina fotográfica que tirou a foto, a data, a lente, e coordenadas GPS. Os metadados de um registo telefónico indicam o tempo da chamada, os seus intervenientes, a quantidade de vezes que aconteceu, ou seja, mesmo sem o conteúdo podemos vir a saber de muita coisa.

Nacionalmente, em 11 de Maio de 2017, as secretas voltaram a ser notícia (quando tal acontece geralmente é por coisas negativas…). Neste dia a comunicação social dava a notícia "*Governo aprova acesso a metadados pelas secretas sob controlo judicial*[126]*"*. Em 2015 o tribunal Constitucional tinha declarado inconstitucional o acesso das "secretas" a dados de comunicações electrónicas: Em 2015, "*o Tribunal Constitucional chumbou uma proposta de lei do governo (com dois votos de vencido), considerando que, segundo a Constituição, a ingerência nas telecomunicações dos cidadãos só poderá ocorrer "no âmbito de um processo criminal". Como os serviços de informações (SIS e SIED) não realizam investigação criminal, o acesso aos dados de tráfego das comunicações estava vedado."*

Em 14 de Julho de 2017, "Lei dos metadados vai ao Tribunal Constitucional"[127]: PCP e Bloco de Esquerda avançaram com pedido de fiscalização do diploma promulgado por Marcelo Rebelo de Sousa.

Em 25 de Agosto de 2017, sai na newsletter de uma empresa de advogados e com o titulo "ACESSO AOS METADADOS PELOS SERVIÇOS DE INFORMAÇÕES DA REPÚBLICA PORTUGUESA" com o seguinte texto: "*a Lei Orgânica n.º 4/2017 vem regular o acesso por parte do Serviço de Informações de Segurança (S.I.S.) e do Serviço de Informações Estratégicas de Defesa (S.I.E.D.) a dados de telecomunicações e internet e que entrará em vigor no próximo dia 30 de Agosto de 2017... "*

Em 14 de Dezembro de 2017, foi devido ao parecer favorável do Tribunal Constitucional referente ao acesso dos metadados pelas forças de segurança, investigação e informações. foi pouco falado tanto nas redes sociais como nos meios de comunicação, mas saiu uma informação importante. Diz o Diário de Notícias: "Constitucional decide constitucionalidade de metadados"[128]. Partidos do PCP, BE e PEV opunham-se quanto à legalidade porque segundo o "*deputado comunista António Filipe, está aquilo que será a violação do artigo 34.º da Constituição, da "inviolabilidade do domicílio e da correspondência. No número n.º 4 deste artigo estabelece-se que "é proibida toda a ingerência das autoridades públicas na correspondência, nas telecomunicações e nos demais meios de comunicação, salvos os casos previstos na lei em matéria de processo criminal".*[129]

Segundo o site metadados.pt são "*informações que acrescem aos dados e que têm como objectivo informar-nos sobre eles para tornar mais fácil a sua organização*[130]*"*, no entanto os metadados por exemplo de um telefone podem revelar algo mais do que a sua "organização"...

[126]https://www.publico.pt/2017/05/11/sociedade/noticia/governo-aprova-acesso-a-metadados-pelas-secretas-sujeito-a-controlo-judicial-1771792

[127]http://www.sabado.pt/portugal/detalhe/lei-dos-metadados-vai-ao-tribunal-constitucional

[128]https://www.dn.pt/portugal/interior/tc-decide-constitucionalidade-de-metadados-8983745.html

[129]https://www.publico.pt/2017/05/11/sociedade/noticia/governo-aprova-acesso-a-metadados-pelas-secretas-sujeito-a-controlo-judicial-1771792

[130]Disponível online em http://www.metadados.pt/oquesaometadados

Por exemplo, os metadados podem dizer quanto tempo pessoa A telefonou para a pessoa B, quantas vezes, só não diz o conteúdo da conversa, mas o resto está lá, até com dados como a possível geolocalização, *software* utilizado, etc…

Esta foi uma grande vitória para todas as forças acima referidas e que lhes poderão dar bom uso, quer de demonstração de prova quer de justificação de aprofundamento da investigação, quer outros. Os metadados são um assunto sério e que podem dizer tanto ou mais que o próprio conteúdo. A sua utilização pode ter efeitos muito nefastos para a democracia. Mas se bem usada, pode permitir "adivinhar perigo e evitá-los."

Os metadados se forem obtidos em fontes abertas são OSINT. O que as forças de segurança e de investigação fazem não são OSINT, mas é compreensível porquê.

2.21.4 Governos

O uso de OSINT pelos Governos é conhecido, tanto pelo seu uso de investigação, como pela prática de contra-informação que os Governos podem fazer junto dos *media*, razão pela qual, é tão necessária a HUMINT com pessoas no terreno para confirmar a veracidade das informações.

Os governos podem ser de natureza totalitária, ou até democrática, no entanto, dada a ocorrência de ataques terroristas nos anos mais recentes, e o impacto (negativo) que estes causam internamente e junto da opinião pública, foram tomadas medidas extra de vigilância. No caso de governos totalitários, o perigo não é tanto de terroristas, mas sim de opositores internos, ou mesmo jornalistas, que comprometam a posição (ou meramente a imagem) das figuras actualmente no poder.

Governos como o chinês, censuram e vigiam fortemente a internet, proibindo tudo o que seja VPN, sites de redes sociais não-chinesas e maior parte dos sítios web.

O governo russo, à semelhança do chinês, mas ainda assim menos rígido, proibiu a presença de redes sociais que não tivessem servidores no seu país. A Rússia em 2015, teve um dos seus representantes das telecomunicações a dizer que se deviam banir as redes VPN e TOR[131]: "*Davydov went on to say that banning anonymising networks would increase user-trust among the Russian people and lead to economic benefits, having described Tor as an 'Anonymous network used primarily to commit crimes".* Já em 2018, baniu a aplicação de conversação cifrada, Telegram[132], na sequência de um caso judicial[133].

Já os governos americanos, inglês, australiano e francês (possivelmente também outros), tentam monitorar todas as comunicações com a desculpa de detectarem e evitarem ataques. Estas monitorizações podem dar lugar a eventos despropositados pois a captura de

[131] https://thestack.com/world/2015/02/11/russia-readying-for-attempt-to-ban-tor-vpns-and-other-anonymising-tools/

[132] "*The Telegram Ban Is Forcing Ordinary Russians to Break the Law (Op-ed*)", disponível online em https://www.themoscowtimes.com/2018/04/24/telegram-ban-is-forcing-ordinary-russians-to-break-the-law-opinion

[133] Telegram foi banido em Moscovo devido a uma acção de tribunal em 13 de Abril de 2018. Roskomnadzor vs Pavel Durov (criador do Telegram) por este não ter fornecido as chaves de cifra do serviço aos serviços russos.

informação é pelo que se sabe, muitas vezes baseado em heurísticas e na captura de palavras e termos soltos.

A localização física do utilizador pode comprometer a sua liberdade de expressão. Neste caso, o uso de ferramentas VPN fidedignas poderão evitar a sua vigilância e decidir se este corre perigo de vida ou não.

2.21.5 *Stalkers* e *Cyberstalkers*

O *stalking* é hoje, felizmente, algo já facilmente reconhecido pela sociedade e sobre o qual existe muita informação. O *cyberstalking* é, no entanto, o mesmo problema, mas de mais difícil reconhecimento e atinge proporções maiores e mais vexatórias, razão pela qual a vítima muitas vezes tem vergonha e por isso não se queixa ou não tem sequer que o *cyberstalking* também é crime.

O Cyberstalking é também a perseguição de uma pessoa, de forma contínua, não desejada e mentalmente agressiva. Geralmente por alguém com quem não tem qualquer relação ou já teve (ex: ex-companheiro). Os motivos são variados, mas geralmente afectivos, e traduzem-se no *Cyberstalking*, por perseguição à vítima através da colocação online de falsas informações, obtenção de informações privadas (nomeadamente moradas, fotos, …), chantagem, monitorização do comportamento da vítima, alteração a dados e sabotagem de equipamentos.

A forma de Cyberstalker confere ao agressor, uma falsa sensação de segurança e impunidade. Por ser na forma virtual, é possível o agressor conseguir agredir a vítima à distância, sem a conhecer, e de uma forma que a vítima não tem qualquer meio de defesa ou de contra-atacar.

A este problema acresce o crescimento do mercado e da procura de software para stalking, chamado de *stalkerware (sobre este tema recomenda-se a leitura do artigo do Jornal Público, "Vigilância e assédio à distância de uma aplicação"[134]). O stalkerware* permite secretamente, aceder, de forma remota, ao conteúdo de telemóveis e equipamentos electrónicos da vitima, permitindo consultar e controlar. Este controlo não é percebido pela vitima.

*A*lguns países já possuem legislação[135] relativa a Cyberstalking, incluindo Portugal, pela APAV[136] que tem um site super completo e com legislação sobre o tema[137], sendo que em Portugal, já é criminalizado desde 2014[138] e no qual se incluiem os casamentos forçados.

2.21.6 Endereço IP

[134]Vigilância e assédio à distância de uma aplicação, disponível online em https://www.publico.pt/2019/01/28/tecnologia/noticia/ciberperseguicao-cyberstalking-stalkerware-1859673, de 2019/01/28, acedido em 2019/06/05

[135] Legislação de combate à Ciberperseguição, disponível em https://pt.wikipedia.org/wiki/Cyberstalking

[136] APAV – associação de apoio à vitima, disponível online em https://www.apav.pt

[137] Levar o Stalking a sério – disponível online em https://apav.pt/stalking/,

[138] Stalking, cyberstalking e casamento forçado vão ser crimes, disponível online em https://observador.pt/2014/09/04/stalking-cyberstalking-e-casamento-forcado-vao-ser-crimes/

Em OSINT, este é um elemento que dificilmente se obtém, no ententanto identifica *à priori* um utilizador e seu local de origem, até mesmo se estiver a usar uma VPN para se esconder.

O IP é um endereço electrónico que identifica um dispositivo tal como um computador, uma impressora ou um telemóvel em redes locais ou públicas, num determinado momento. Cada dispositivo ligado à rede possui um IP (e um *MAC address*[139] único). O ip é uma forma de identificação para as máquinas se comunicarem na Internet e é facilmente identificável quando acedemos a um sítio web ou utilizamos um qualquer serviço. Alguns endereços IP estão associados a organizações e/ou países, pelo que é bastante comum, o bloqueio de algumas gamas/ranges de ip para bloquear ips tidos como maliciosos. Existem serviços na internet que através do endereço IP que lhes forneçamos, nos informam da cidade e localização aproximada do endereço IP. Exemplo de endereço IP: 192.168.1.1

2.21.7 Geolocalização

A geolocalização é como diz o nome, a localização geográfica. De enorme utilidade e fornecida por diversas ferramentas, hoje em dia todos as têm, quer seja pelo uso de computadores, máquinas fotográficas, sistemas GPS, etc.

O que víamos há uns anos atrás em filmes de agentes secretos, está presente hoje em dia de uma forma prática, barata e simples. Através da geolocalização, os motores de busca podem facilitar a entrega de publicidade para os seus clientes, sabendo de antemão onde estes estão e o que podem ter necessidade de saber/comprar. Em telemóveis de idosos, pode ser uma utilidade pois o típico botão de emergência pode enviar um sinal de alerta e a localização GPS para os números gravados, por exemplo da família e posto da polícia mais perto.

Ao utilizarmos tecnologia, deixamos pequenas marcas que podem ser analisadas e rastreadas até à origem. Já foram referidas algumas. No caso da geolocalização, a menos que estejamos a usar uma VPN, isto é garantido, o endereço IP fornece estes dados.

As máquinas podem ser localizadas através do seu IP, *mac address, RFID*, coordenadas *GPS* do seu telemóvel, aplicações do telemóvel, metadados de fotos, informações dos operadores de telecomunicações (triangulação de dados), identificações nossas ou de terceiros nas redes sociais ou em sites pessoais, entre outros.

Nas redes sociais, esta localização permite encontrar amigos/as. Permite também que nos sites em que as fotos não sejam devidamente "anonimizadas", seja possível saber onde foi tirada, e se a foto foi tirada à frente de casa ou tem elementos que identifiquem a localização (carros, matrículas, restaurantes, …) então o investigador/*hacker*/atacante/polícia, já sabe "tudo". Permite também saber onde está alojado o nosso site/serviços.

[139] *MAC address* – MAC significa *Media Access Control*, e é uma identificação física associada à interface de comunicação (tipicamente uma placa de rede), que está a permitir que o equipamento se ligue à rede/internet. Não podem existir dois endereços MAC iguais (a não ser que seja forjado propositadamente)

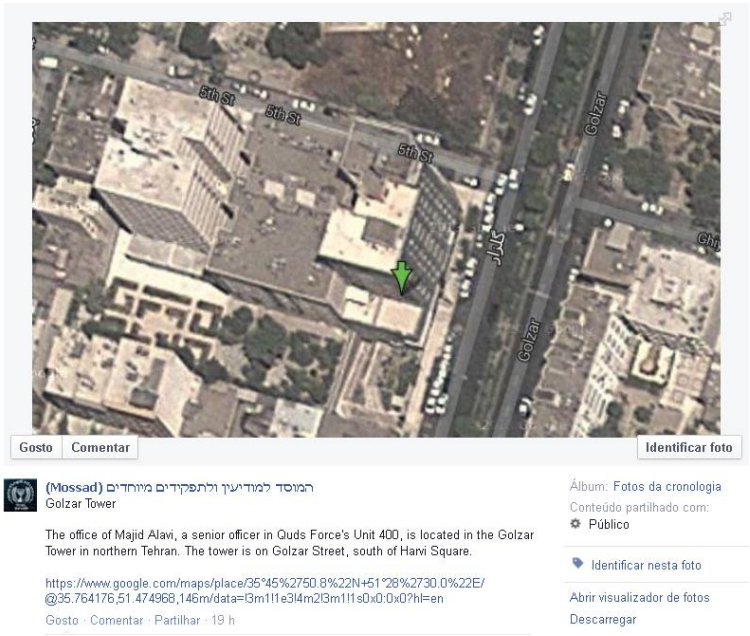


Figura 12 - Obter dados GPS, imagens e id de pessoas através de blogues e redes sociais

O projecto iKNOW, possui ferramentas para retirar de imagens, conteúdos EXIF, entre eles, as coordenadas do local onde foi tirada a foto (se esta tiver estas informações).

2.22 Fugir à detecção OSINT no ciberespaço

Assim como a OSINT pode ser utilizada para o "bem", outros a utilizarão para propósitos menos honestos e até ilegais. A primeira premissa para não sermos detectados e identificados com facilidade, é não colocarmos na Internet, elementos pessoais que nos possam identificar. A utilização de redes sociais, tão comum hoje em dia, deve ser de evitar, ou pelo menos com fotos e outros elementos únicos que nos identifiquem e nos coloquem em locais e identificados no tempo.

Há depois outras técnicas que se tentarão aqui sumariar de forma a não ficarmos demasiado expostos nem a atrair demasiada atenção.

2.22.1 Integração e discrição – "Ser como eles"

Como na maior parte das coisas, a melhor forma de se esconder é dar o menos possível nas vistas, ser discreto, falar (no caso do online é escrever) pouco. Assemelharmo-nos ao meio que nos rodeia, ser como as pessoas locais e sua cultura. Isto é especialmente direccionado para quem está em território hostil, sem liberdade de expressão, ou semelhante.

- Assemelharmo-nos e agirmos como eles. Em caso de discórdia, não opinar (especialmente válido nas redes sociais);
- Não deixar padrões (não intencionais) ou fora do comum;
- Utilizar redes, *software* e recursos isolados para análise de *malware;*

- Em redes wifi que não haja "total" confiança, não usar;
- Como teste, pedir a outros que nos avaliem como parecemos perante os outros e perante o nosso alvo;
- Utilizar perfis falsos para acções que possam relacionar o nosso perfil com o perfil de outros elementos ou grupos com os quais não desejamos ser relacionados (pode esta atitude fugir ao OSINT? Pode. Espionagem? Depende. Não estamos a ultrapassar barreiras de segurança e qualquer pessoa pode aceder à informação mediante aceitação no grupo, mas será ético? Talvez não, mas é OSINT.);

2.22.2 Métodos e formas de anonimização

Novamente direccionado para quem não quer aparecer nos "alarmes" OSINT ou para quem está situado em território hostil ou sem liberdade de expressão, ou é jornalista, ou semelhante.
Além de aparentar sermos iguais, as nossas acções no mundo online têm de permitir executar o que se pretende sem com isso, gerar alertas. Pode acontecer que o nosso equipamento tenha de analisado por terceiros, e não pode ter indícios de comunicações que levantem suspeitas.
Seguem-se alguns procedimentos de segurança e anonimização:

- Computador comprado a dinheiro num local não filmado;
- Sistema Operativo não registado (Windows "pirata" ou usado em modo "trial" ou em alternativa, utilizar um sistema operativo Linux ou *NIX (OpenBSD FreeBSD, …)), com preferência para sistemas operativos Linux direccionados para a privacidade (Tails, Qubes, Whonix, Subgraph OS);
- Falsificação do *MAC address* antes de ligar à Internet;
- Utilização de ligação VPN e/ou rede TOR;
- Utilização de rede wireless gratuita em *sítio público* em que não seja necessário *login;*
- O dito sítio publico deve ter sido visitado anteriormente para verificar se não existem câmaras de vídeo e afins;
- Usar um perfil falso online com passes diferentes das reais;

As técnicas abaixo serão mais uma forma de tentar escapar aos radares, e que seguem a mesma lógica da descrição e da integração acima referidos.

Não atribuição (não parecer com ninguém em particular)

- Misturar meios técnicos e sociais/humano de mistura
- A localização é o segundo item mais procurado a seguir à identidade.
- Quando em Roma... Sê romano.
- Que sites de redes sociais?
- Que salas de conversação?

Várias identidades... Nenhuma identidade

- Não fazer *login* em sites, tendo de ser (utilizar identidades/perfis aleatórios e falsos)
- Usar *https* nas pesquisas e nas salas de conversação, redes sociais, etc. Evitar usar seja o que for via web e que não seja utilizado *HTTPS*
- Utilizar IP's aleatórios (usar a internet fixa de casa, onde o ip geralmente demora dias para mudar não serve)

- Usar a rede TOR ou VPN (redes VPN boas e rápidas são pagas, mas tem-se uma ligação que é prometida anónima – não esquecer de pagar esta ligação VPN de forma anónima)
- Não aceder a emails nem a sites que seja necessário autenticação (*logins*...)
- Usar um navegador que anonimize navegação e não corra *plugins* (por defeito, apenas os autorizados manualmente)

Minimizar padrões de reconhecimento

- Criar e utilizar identidades/perfis aleatórios e falsos
- Estar pouco tempo, seja no ciberespaço com o mesmo IP (com a rede TOR é possível forçar a mudança automaticamente. Com a rede VPN também podemos fazer esta mudança de ips de forma simples e rápida, escolhendo até o país de saída).
- Possibilidade de usar máquinas virtuais com diferentes S.O. juntamente com os perfis falsos
- Apagar histórico de navegação e *cookies*
- Usar um navegador que anonimize navegação e não corra *plugins*

2.22.3 Teste à presença online através do navegador

O sítio web panopticlick[140] (da reputada *Electronic Frontier Foundation EFF*[141]) é uma excelente ferramenta para testar se estamos a utilizar demasiados elementos identificadores quando acedemos a um sítio web com o nosso navegador. A imagem abaixo mostra todos os dados passíveis de ser obtidos através da visita a um sítio web. É assustadora, a quantidade de dados e a forma como estes podem ser correlacionados para identificar uma pessoa, mesmo que utilize outro endereço ip. Utilizar um browser como o TOR e um sistema operativo que funcione numa *pen* USB, rejeitar cookies e *plugins*, pode ajudar a minimizar a nossa pegada digital. Fica um resultado do teste efectuado como (comum e mau exemplo):

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	6.29	78.44	Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:65.0) Gecko/20100101 Firefox/65.0
HTTP_ACCEPT Headers	10.33	1286.34	text/html, */*; q=0.01 gzip, deflate, br pt-PT,pt;q=0.8,en;q=0.5,en-US;q=0.3
Browser Plugin Details	0.86	1.81	undefined
Time Zone	2.16	4.48	0
Screen Size and Color Depth	2.55	5.85	1920x1080x24
System Fonts	5.21	36.94	Arial, Arial Black, Arial Narrow, Arial Rounded MT Bold, Arial Unicode MS, Book Antiqua, Bookman Old Style, Calibri, Cambria, Cambria Math, Century, Century Gothic, Century Schoolbook, Comic Sans MS, Consolas, Courier, Courier New, Garamond, Georgia, Helvetica, Impact, Lucida Bright, Lucida Calligraphy, Lucida Console, Lucida Fax, Lucida Handwriting, Lucida Sans, Lucida Sans Typewriter, Lucida Sans Unicode, Microsoft Sans Serif, Monotype Corsiva, MS Gothic, MS Outlook, MS PGothic, MS Reference Sans Serif, MS Sans Serif, MS Serif, Palatino Linotype, Segoe Print, Segoe Script, Segoe UI, Segoe UI Light, Segoe UI Semibold, Segoe UI Symbol, Tahoma, Times, Times New Roman, Trebuchet MS, Verdana, Wingdings, Wingdings 2, Wingdings 3 (via javascript)
Are Cookies Enabled?	0.19	1.14	Yes
Limited supercookie test	0.27	1.2	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No
Hash of canvas fingerprint	7.55	187.89	2287994a42ffa32545d5242399db51f5
Hash of WebGL fingerprint	4.23	18.77	7a06e495f624fb535c8b5acac4c78409
DNT Header Enabled?	0.76	1.69	True
Language	9.07	536.29	pt-PT
Platform	1.29	2.45	Win32
Touch Support	0.55	1.47	Max touchpoints: 0; TouchEvent supported: false; onTouchStart supported: false

Figura 13 - Fingerprint da nossa identidade, usando unicamente o navegador web

[140] https://panopticlick.eff.org/
[141] https://www.eff.org/

Além da imagem, o sítio fornece-nos informação sobre o que considera ser a nossa pegada. No caso acima, é-nos dito que **o nosso navegador aparenta ser único em cerca de 184 mil navegadores**.

*"Your browser fingerprint **appears to be unique** among the 183,947 tested in the past 45 days. Currently, we estimate that your browser has a fingerprint that conveys **at least 17.49 bits of identifying information."***

2.23 Conclusões

O OSINT não vem tirar lugar a nenhuma forma de inteligência nem tirar ou dar protagonismo a nenhuma instituição em concreto. Sendo fontes abertas, estão ao recurso de todas as forças de segurança e de todos os cidadãos. Tem sido observado que o OSINT é cada vez mais utilizado e valorizado. Foi observado no sítio web da Amazon que praticamente todos os livros de OSINT têm data recente(ver anexo *Bibliografia especializada em OSINT*).

Tanto o SIS como o SIED, têm a obrigação de comunicar entre si, dados (estratégicos[142]) que possam contribuir para as suas funções e/ou para entidades nacionais que possam ser afectadas por ameaças quer nacionais quer transnacionais. A informação OSINT facilita a partilha e a confirmação de factos.

Ou até para entidades das quais Portugal faz parte, como por exemplo a NATO ou uma das estruturas europeias da União Europeia, entre outros. Também a Polícia Judiciária deve ser informada por estes dois serviços. Se a PJ souber de algo que possa acontecer, também deve informar reciprocamente. No entanto, só a Polícia Judiciária pode actuar criminalmente.

José Vegar, escritor, em entrevista ao programa televisivo Inferno sobre o tema dos serviços de informações: "…*se não se perceber a utilidade dos serviços de informações, eles não podem trabalhar. E nós precisamos deles como cidadãos*".

GNR: Em 2017, foi aprovado o projecto PT/2016/FSI/109, que vem criar um "Centro OSINT" na GNR e que segundo a própria informação[143] [144] da Secretaria Geral do MAI, "*visa a criação de uma Unidade OSINT que permitirá à GNR implementar e gerir um sistema alarmístico suportado pela inteligência resultante do ciclo de produção de informações, onde as necessidades de informação oportuna serão geradas e definidas, recolhidas, integradas, analisadas, avaliadas e disseminadas. Com a criação desta unidade a GNR pretende dar continuidade e suporte à prevenção e combate ao crime, garantindo a interoperabilidade e a continuidade de sistemas e*

[142] ver referência a informações estratégicas em "2.3 Geração e classificação de informações pela Defesa e forças militares"

[143]Documento com objectivos, projecto e custos do "Centro OSINT da Guarda" disponível em http://www.gnr.pt/ficheiros/seguranca_interna/3.pdf

[144]Fundo de Segurança Interna financia Unidade OSINT da Guarda Nacional Republicana - https://www.sg.mai.gov.pt/Noticias/Paginas/Fundo-de-Seguran%C3%A7a-Interna-financia-Unidade-OSINT-da-Guarda-Nacional-Republicana.aspx

processos, de forma a incrementar a cooperação e a troca de informação entre os diversos organismos e a fortalecer a capacidade de análise de informações…"

CAPÍTULO III - Sistema Proposto e Hipótese de Investigação

3.1 Objectivos, potencialidades e diferenciação

A ideia da criação da plataforma iKNOW deveu-se ao gosto pelo tema, pela crescente importância da OSINT num tempo que todos partilham tudo, mas também, devido à necessidade do autor(*não foi encontrado nada que pudesse resolver os desafios abaixo*) pela:

1. automação da captura de informações *(mesmo em sites onde os seus autores não pretendem permitir a partilha, apenas leitura*), evitando ao máximo informações desnecessárias (ex: publicidade). Informações como palavras-chave, obtenção de conteúdos de sites, imagens, metadados, assim como conteúdos de sítios da chamada *dark web*;
2. automação do processamento da informação recolhida, através da introdução destes dados numa base de dados, categorização da informação, entre outros;
3. análise estruturada da informação (informações que coloquem a informação num sítio com ambiente e tempo próprios). Datas, autores, categorias, tipo de informação, etc;
4. análise, complementarização, e geração de relatórios OSINT.

Tudo a partir de fontes de informação abertas, quer seja os media tradicionais, como também sites de informação, sites técnicos, sites com palavras-chave, sites de capas dos jornais diários, sites da policia (várias), entre muitos outros. O relatório, que é feito automaticamente e compilado numa base constante, foi pensado para ser distribuído (ex: no final da semana). Pretende-se ter ao longo do dia, informações diversas e que nos possam ajudar a nós e às nossas organizações, SOCs, CSIRTs, a precaver-se (ter a informação que um site vai ser atacado, pode não evitar o ataque, mas impedir que tenha consequências sérias) e a tomar decisões.

O iKNOW é uma ferramenta desenvolvida para actuar de forma distribuída, com um ponto central (servidor web e base de dados centralizados) onde são colectados os dados e de onde é depois gerado o relatório e métricas pretendidas. O utilizador deve fazer um registo inicial a partir do qual, pode navegar no site, efectuar diversas "actividades" e inserir as operações que pretende que o iKNOW efectue (*ver apêndice 6.3 para ver funcionamento*). O utilizador pode inclusive, registar e pôr ao serviço, os seus próprios equipamentos (*Raspberry*).

Os dados podem ser tão diversos como imagens, texto, coordenadas GPS, hipertexto, ficheiros de um determinado tipo, entre outros.

Durante a escrita desta dissertação, foram encontrados diversos *sites* que pretendiam chamar OSINT à cópia massiva dos conteúdos de sites de x em x tempo. Ao contrário destes, o iKNOW percorre o *site*, procura pelos termos e continua a procurar nos níveis pretendidos, apenas guardando informação se encontrar realmente alguma coisa. Pode, no entanto, ser acrescentado ao código, a funcionalidade de gravar tudo logo que detecte uma alteração. Isto

já foi feito, recorrendo a funções de *hash* [145] do *site* de tempo em tempo e verificando se este *hash* sofria alguma alteração.

Ainda durante a escrita, foi repensada a ferramenta para responder a uma série de desafios, que passavam pela mais rápida e precisa introdução de informações, edição das mesmas e posterior emissão de relatórios.

3.2 Método OSINT – O ciclo de produção de informações

A produção de *intelligence* OSINT exige um método, sob pena de nos perdermos em tanta informação ou não gerarmos algo realmente novo. É aceite universalmente o ciclo de produção composta por cinco fases;

Figura 14 - Ciclo de produção de informações

O planeamento é a linha mestra que deve definir qual o objectivo que se pretende, tipicamente definido pelas estruturas superiores da organização.

A recolha da informação é pensada pelos técnicos e analistas, para ser feito de forma mais granular. Não se pretende tudo, apenas o que possa interessar. É importante aqui ter dados com que trabalhar. Neste aspecto, o OSINT pode ser a melhor ferramenta para se trabalhar.

Análise, processamento e produção de dados: com OSINT é comum ter mais dados que o que é possível analisar (a menos que trabalhemos num serviço de informações e mesmo assim..), pelo que se tem de armazenar de alguma forma todos esses dados e só depois, usar um método para os filtrar e analisar, de forma que no final, quando se produz informação, esta ainda possa ser

[145] *Hash* é um processo matemático, unidireccional, que impossibilita descobrir o conteúdo original a partir do seu valor. Este valor muda se for alterado qualquer coisa (acrescentado ou retirado) da mensagem. O código *hash* varia conforme o método (algoritmo) utilizado, assim como o seu comprimento. Exemplo de *hash* com o algoritmo SHA1 para "olá mundo": 47998549b0b97cc9aab0bcd64ce036fc3edc4754.
Site onde é possível testar hash: https://www.fileformat.info/tool/hash.htm

utilizada sem se ter perdido tempo a mais, nem se tenha perdido a informação útil no meio de tanta outra.

Disseminação: as conclusões retiradas e a forma como são expostas devem ser adaptadas a quem as vai receber. Isto pode fazer a diferença entre uma informação boa e utilizada e uma informação que podia ser boa, mas não foi compreendida.

O *feedback* é um passo quase facultativo. É importante na medida que pode gerar nova necessidade de planeamento e recolha ou a melhoria dos processos.

Estes passos acima descritos também foram observados em 4 passos: Descoberta, Selecção, Análise e filtragem, Resultados e entrega.

Exemplo: obter informações no *PasteBin*[146] de todos os "*posts*" que lá forem inseridos, que contenham no seu interior, o termo "*daesh*".

Como se pretende tudo o que tenha "*daesh*", independentemente se no inicio ou fim, podemos recorrer a uma expressão regular "*regex*"[147] que permita esta condição de procura e depois de um *crawler* que varra o site em causa, em busca de todas as publicações, para procurar no seu interior a nossa expressão. O que for encontrado vai ser depois enviado para uma base de dados, juntamente com o endereço da página, a data/hora e o conteúdo completo (não vá ele ser apagado...). Depois no final do dia, a aplicação geraria um relatório com os endereços encontrados, autor, data/hora e conteúdo. Esta técnica foi utilizada em alguns scripts iKNOW para efectuar pesquisas.

Na figura 15 é mostrado um esboço do processo descrito. Utilizando uma pesquisa para o termo *daesh*, ilustra-se o funcionamento da ferramenta iKNOW e passos percorridos.

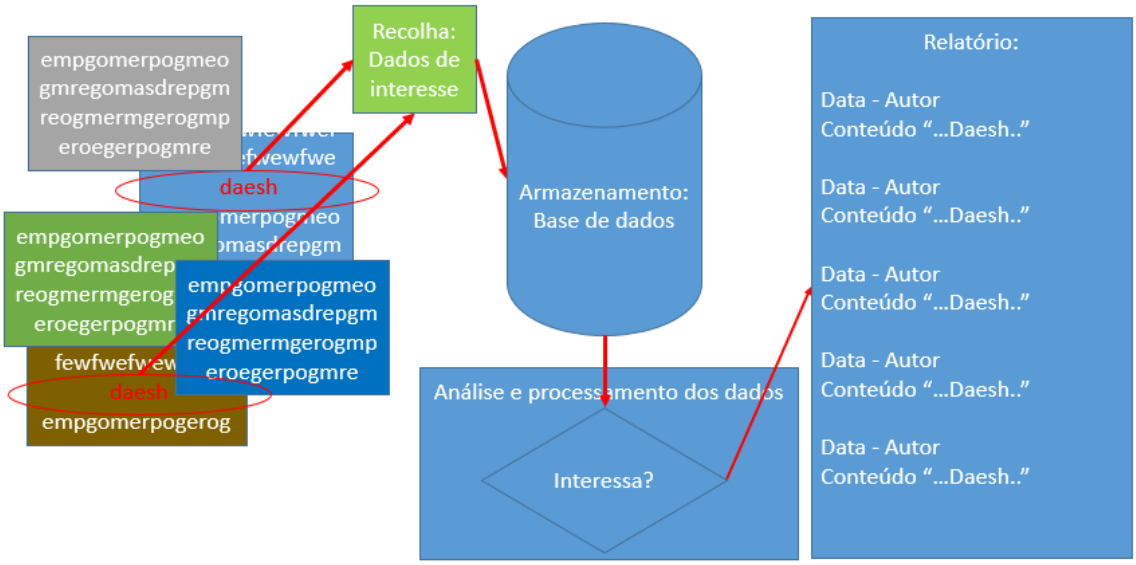


Figura 15 - Prático: ciclo de produção de informações

[146]Sítio Web de partilha anónima de textos, dizeres, projectos, informações variadas. Disponível em www.pastebin.com, consultado nesta data em 5 de dezembro de 2018

[147] Regex – expressão regular, ou seja, um texto formatado de maneira especial e que é interpretado, permitindo que se façam pesquisas em textos, para encontrar padrões. Exemplo: dir *.txt na linha de comandos seria *\.txt

3.3 Hipótese de Investigação – O que se pretende

Relativamente à investigação, ao estado de arte, e aos objectivos, houve essencialmente os pontos abaixo, já anteriormente referidos, e que motivaram os objectivos, a escolha dos meios, as necessidades reais e o funcionamento:

1. **Necessidade pessoal**: relatórios OSINT mais rápidos (menos onerosos em tempo e recursos humanos, nomeadamente do autor). Os relatórios pretendem antever acontecimentos nefastos para a instituição – *ver ponto 3.3.1 para mais detalhes*.
2. **Necessidade institucional/colectiva:** resumo curto mas completo, possibilidade de se ter acesso a uma informação mais longa e/ou a original, a informação tem de ter valor para a pessoa que a vai ler ou informações de relevo sobre a instituição na "comunidade", comunicação social, etc. A informação tem de ser actual.
3. **Ajudar Serviços de Informações e forças de segurança**: localização de equipamento furtado, cartões de crédito furtados ou clonados e/ou sua venda na internet, termos relacionados com criminalidade e terrorismo. Localização de equipamentos e termos na *dark web*, pessoas desaparecidas, termos que poderão causar danos a pessoas ou à sua imagem, assim como de instituições no nosso país. *Ver também ponto 3.3.2*
4. **Custo reduzido, carácter distribuído e âmbito didáctico:** pretende-se que seja uma solução barata e que opere de forma distribuída. Como se trata de uma ferramenta no âmbito de mestrado, a prova de conceito e até de produção, recaiu no uso de equipamentos electrónicos baratos como o mini-computador *raspberry,* sensores e colectores eléctricos/electrónicos. Ver ponto 3.3.3

Pergunta de partida para a ferramenta iKNOW e dissertação de mestrado: *"Será possível a construção de um sistema informático automatizado que:*

1. *permita a construção de relatórios OSINT, com menos tempo, menos esforço, mesma garantia de fiabilidade, maior rapidez de geração, com envio em segurança para quem se pretende e com uma forma de "customizar" a informação pretendida? (Segurança e Integridade);*
2. *que obtenha automaticamente conteúdos (texto de notícias e imagens de capas de jornais);*
3. *E QUE consiga "varrer" sites e localizar informação, classificando-a para se perder menos tempo em sítios web e na sua análise inicial? De forma distribuída para que o "crawl" dos sítios, no caso de ser bloqueada, continue a funcionar? (Granularidade da solução e tem termos de segurança, assegurar Disponibilidade);*
4. *E QUE tudo isto possa ser implementado em sistemas informáticos baratos, sem grande poder de processamento como os Raspberry?;*

Extras:

5. *Que tenha implementado um sistema multi-utilizador, diferenciando pessoas e entidades? (Segurança);*

6. *Que possua ferramentas de suporte que ajudem e sejam úteis ao utilizador, incluindo estatísticas de utilização, métricas, sistema de tickets, entre outros."*

O iKNOW pretende responder a tudo isto. O que se pretende/u é/foi sem dúvida bastante ambicioso e não existe nada assim (*que se conheça ou se tenha encontrado*) na actualidade.

3.3.1 Relatórios OSINT

Nos locais onde o autor tem trabalhado, na área da segurança informática, tem-se dedicado à elaboração de relatórios OSINT (ver relatórios OSINT). Mesmo quando estes ainda não existiam ou não eram prática na instituição.

Estes relatórios, pretendem dar o "panorama" actual da sociedade em relação à instituição e/ou Governo, etc., dependendo do âmbito de serviço. Por exemplo, num determinado organismo governamental, o foco dos relatórios era na antevisão de manifestações, ataques informáticos, nível de empatia sentido pela população para com a instituição, forças policiais, previsão de ocorrências, etc.

A pesquisa para estes relatórios é totalmente feita com recurso a OSINT. Capas de jornais, revistas da actualidade, televisão, e Internet, sobretudo redes sociais e sites de tecnologia. Tudo de forma manual.

Por exemplo, foi possível por diversas vezes, antever ataques informáticos com bastante precisão e alertar para tal. Tal como antever manifestações e/ou alertar para sites web que continham problemas de segurança antes de serem aproveitados.

Estes relatórios não existiam, mas a partir do momento que começaram a ser apresentados, ganharam um carácter de importância. A utilidade não se resume apenas à instituição, pois se a informação for de interesse, poderá ser divulgada para outras instituições do mesmo ramo ou para os serviços de informação ou forças de segurança. Tudo OSINT. Nunca usando de espionagem ou interferindo com sistemas.

O trabalho de construção de um relatório OSINT não é difícil, mas é moroso e necessita de alguma análise conjunta de várias fontes, eliminando à cabeça *fake news* e outros "engodos". Há muita informação, mas há que percorrê-la toda para selecionar o que interessa. Quando existe... Há muitos dias sem informação de interesse (algo realmente relevante). Os relatórios também abarcam notícias de redes sociais, e sítios web de segurança informática de outros países, aliados e não-aliados…

Os relatórios, por terem interesses diferentes, obrigaram à criação de um sistema multi-utilizador, em que cada um pode fazer relatórios diferentes. As noticias (para já) pretendem-se comuns a todos.

3.3.2 Obtenção de informações, para serviços de informações e investigação

Outra potencialidade que se gostaria de ver na aplicação iKNOW, tem como objectivo auxiliar de alguma forma os serviços de informações e os órgãos de investigação.

Uma das motivações prende-se com um problema pessoal: quando assaltaram o carro do autor, levaram alguns pertences. Tendo sido feita a queixa nos órgãos próprios e dados os números de

série e o modelo do que era possível (portátil, discos, banda larga móvel, *router*, …) entre também mochilas e material que ia para fim de semana. O objectivo deste ponto, é a localização de material roubado em fórums e sites conhecidos de vendas online, pesquisando pela marca, modelo, e outros items que possam ajudar à identificação posterior. Um dos pontos sugeridos com alguém de um órgão de investigação, seria a localização de cartões de créditos furtados. Poder localizar padrões de cartões na rede TOR, também seria interessante.

3.3.3 Custo, segurança e funcionamento distribuído

Sendo este um curso vocacionado para a segurança, queremos implementar alguns extras de segurança, como por exemplo, autenticações seguras, tratamento de *inputs*, envio de correio de forma cifrada, comunicações cifradas, rede TOR, entre outros. Sendo uma ferramenta que se propõe operar de forma distribuída, tem de ser barato, consumir pouca electricidade e ser fácil de montar e manter.

Custo: foi escolhido o *Raspberry Pi*[148], como cliente e servidor da plataforma, embora possa ser utilizado como servidor um sistema informático normal com Windows, Linux, …
O *Raspberry Pi* tem as dimensões de um cartão de crédito, podendo ser transportado para qualquer sitio, por qualquer pessoa. Tem um custo reduzido (comparativamente a um computador), aproximadamente de 35-40 euros e tem poder de processamento suficiente para poder ter um sistema operativo, poder servir páginas web, fazer de proxy entre tantos outros. A escolha no *Raspberry*, deveu-se essencialmente ao seu baixo custo, e devido a ser tão pequeno e tão facilmente transportável, poder ser uma plataforma portátil. Ter uma porta de rede, portas USB e um sistema Linux foram tudo extras muito bem-vindos.

Os pormenores sobre a execução, o que foi feito, etc. estão no capítulo de implementação.

Segurança: a plataforma opera sem necessidade de o cliente ter monitor. É simples de instalar e por defeito, no *script* fornecido, são logo fechadas as portas do equipamento, é alterada a palavra-passe, e instalado tudo automaticamente (inclusive TOR, certificados web). Apenas funciona SSH (no porto não-padrão 56789), HTTPS com certificados e pouco mais. A firewall é fechada para reduzir a superfície de ataque ao mínimo.

O servidor é instalado de forma diferente, mas contém também as mesmas condições de segurança e outras mais, devido à plataforma web, que vai receber autenticações e possuir formulários. Dado que o servidor pode executar comandos de sistema a partir de páginas web, devido às permissões intrínsecas que lhe foram dadas, foi desabilitada a possibilidade para que apenas um utilizador possa enviar ficheiros directamente (*uploads*).

Nos testes efectuados, depois da autenticação efectuada foi possível adulterar conteúdos pelo que o registo de novos utilizadores teve de ser fechado. É um ponto a melhorar.

Funcionamento distribuído: A Universidade de Cambridge, explica na sua página web[149], que a

[148] O Raspberry Pi é um mini-computador pessoal. Hoje em dia muito conhecido e utilizado por todos, inclusive em escolas para colocar os mais novos a dar os primeiros passos na programação.
[149] "What is Distributed Computing?" Universidade de Cambridge. Disponível online em https://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/distributed-computing/

computação distribuída envolve partir um problema computacional em diversas partes e tarefas paralelas, a serem completadas por dois ou mais computadores numa rede. Permitindo assim resolver problemas que envolvam um grande número de dados ou de iterações.

Pretende-se com o trabalho realizado que o iKNOW, não seja um super-computador, mas que divida as suas tarefas com todos os seus obreiros. O servidor não deve fazer pedidos, mas sim colocar estes pedidos em listas para que os obreiros peçam, um de cada vez, executem, e devolvam o resultado. Ou seja, uma tarefa que pode ser obter um site, pode ser feito por vários sistemas. O objectivo, como referido anteriormente é que nenhum sistema cliente ou servidor seja bloqueado por excesso de pedidos ou comportamentos anómalos.

3.3.4 Técnicas de obtenção de dados sem "gerar alarmes"

Quando uma pessoa faz pesquisas online por informação ou analisa dados, demora tempo. Na pesquisa lê e vai seleccionando apenas o que quer, vai possivelmente saltar alguns dados "menores" e focar outros, aqueles que acha ser importantes. Sabe o que quer e vai escolhendo sem clicar em tudo.

Automaticamente, as buscas não têm o factor inteligente. Geram um volume bastante elevado de dados e não possuem um critério de selecção, o que pode levar a entrar em sítios onde uma pessoa não entraria (hiperligações para sites maliciosos, cliques em hiperligações ocultas, *download* e cliques em publicidade...).

A solução passa pela utilização de métricas, de modo a dissimular a máquina e os automatismos: atrasando o número de chamadas/cliques/*get's*/. por alvo e por fonte em cada minuto/hora/dia, podemos simular um humano. Praticamente todas as *firewalls* de perímetro e de aplicação (*WAF*) conseguem detectar *bots*. Esta solução rodeia isto.
Técnicas: Simulação de um utilizador humano.

- Um utilizador por norma, não usa o CURL ou outra ferramenta do género para navegar num site, portanto, temos de assemelhar sermos "apenas mais um". Devemos colocar um *user-agent* conhecido, por exemplo: *Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.2 (KHTML, like Gecko) Chrome/22.0.1216.0 Safari/537.2*
- Pedidos limitados, e espaçados aleatoriamente no tempo: entrar no site, descarregar, demorar x segundos até abrir uma nova ligação. Esta técnica não sobrecarrega o alvo com pedidos, assim além de "ético", não geramos alarmes DoS/DDoS nem somos bloqueados. 1º pedido aleatório entre 5 e 10 segundos, 2º pedido *idem* e por aí.
- Várias origens/máquinas a fazer pedidos, podem desviar a atenção de uma máquina em particular.
- Simulação de cliques de pessoas. Não clicar em hiperligações dúbias, com outros assuntos ou publicidade, assim como não clicar em hiperligações escondidas (possíveis *honeypots*).
- Definir onde estão os parágrafos de texto que nos interessam, seleccionar o título, autor e datas. Isto deve ser adaptado se os primeiros resultados não chegarem bem (formatação, texto a mais, língua estranha com caracteres não aceites pelo interpretador/*parser* ou de inserção difícil na base de dados, entre outros, …).

As técnicas acima pretendem ser colocadas em prática nos equipamentos e *scripts* do lado

cliente. O exemplo abaixo é a autenticação numa base de dados a partir da sua interface web.

Exemplo prático de simular e automatizar autenticação num sítio web

Alerta: o que abaixo é mostrado pode ser utilizado não só para autenticar automaticamente num determinado sítio web, mas para testar por força bruta, credenciais até acertar (neste caso já não seria OSINT porque estaríamos a ter acesso a informação que não temos autorização/credenciais).

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

import mechanize
import cookielib  #necessário para guardarmos os cookies recebidos futuramente pelo sítio web
br = mechanize.Browser()

cj = cookielib.LWPCookieJar()
br.set_cookiejar(cj)

#escapar a robots simulando comportamentos, e usando apenas aqueles que queremos
br.set_handle_equiv(True)
br.set_handle_gzip(True)
br.set_handle_redirect(True)
br.set_handle_referer(True)
br.set_handle_robots(False)
br.set_handle_refresh(mechanize._http.HTTPRefreshProcessor(), max_time=1)

# falsificacao de user-agent para escapar a sistemas que detectam aplicações e ataques desconhecidos ou maliciosos
br.addheaders = [('User-agent', 'Mozilla/30.0 (X11; U; Linux i686; en-US; rv:1.9.0.1) Gecko/2008071615 Ubuntu/14.0.1 Firefox/30.0.1')]

#Testar éticamente com o nosso próprio servidor....
r = br.open('http://localhost/phpmyadmin')
html = r.read()

# mostrar o titulo e os headers
print br.title()
print r.info()

# mostrar se o site tem formulários que possamos usar. Serve isto apenas para testes e depois implementar
print "MOSTRAR FORMULARIOS:"
xx=0
for f in br.forms():
   print xx, ": ",f
   xx=xx+1

# seleccionar o primeiro formulário/caixa, chamado de index zero. Neste caso queremos o segundo item
```

```
br.select_form(nr=1)

#aqui preenchemos o formulário com os valores abaixo e fazemos submit/enviar para entrar
br.form['pma_username']='root'
br.form['pma_password']='anossapasse'
br.submit()
print br.response().read()

#se tudo deu certo (e a passe e utilizador estiverem correctos) devemos ter aqui muito muito texto, que é a interface web.
```

O exemplo fornecido funciona em páginas web simples, fóruns com autenticação, bases de dados com interface html, entre tantos outros. O mesmo tipo de simulação é utilizado para percorrer hiperligações, simulando neste caso também o tempo que um utilizador demoraria até clicar numa hiperligação e depois outra e outra.
As bibliotecas *mechanize* e *cookielib* são muito úteis permitindo automatizar a guarda dos cookies e algumas acções. Outra biblioteca deveras utilizada foi a *requests*, que com pedidos simples, nos permite obter diferentes respostas, úteis não só para obter dados mas também para nos indicar o estado do servidor, páginas e erros, entre outros.

CAPÍTULO IV - Implementação e casos de estudo

4.1 Implementação

SE a informação existir, com as técnicas certas, as informações certas, e alguma sorte é possível obter tudo sobre uma pessoa, instituição ou situação, com base em OSINT, bastando que para tal, alguém ou algo (sistemas informáticos com inteligência artificial, câmaras de vigilância, bases de dados, …) injecte na Internet ou nos meios de comunicação, informações. O correlacionamento e a correcta digestão das informações gera a tal "intelligence" que torna útil o que se capturou. O iKNOW pretende ser uma ferramenta que ajuda na obtenção destes elementos. Será o analista a fazer depois a correlação. O iKNOW não tem Inteligência Artificial ou mecanismos de selecção. Contém, no entanto, algumas habilidades de análise de texto, expressões e "analytics" que nos podem dar algum grau de certeza relativamente à informação ser boa ou não (ex: se um sítio web tiver escrito a palavra bomba, Portugal, e alguns termos relacionados com Estado Islâmico, estaremos com certeza diante um sítio web de interesse e que devemos verificar manualmente para inclusão ou não nos relatórios pretendidos).

Seguem-se alguns casos práticos em que o OSINT obtido é de grande valor e nos permite obter informação útil para nós e para os outros, o projecto iKNOW, sua apresentação, evolução e utilidade. A ferramenta apresentada mutou várias versões no tempo que demorou a ser desenvolvida e o resultado final ficou ligeiramente diferente do inicialmente pretendido e feito.

4.2 Caso real de investigação – Descoberta total da identidade do hacker

O caso que a seguir se apresenta não utilizou a ferramenta iKNOW, mas pretende demonstrar a potencialidade da OSINT e da sua análise e aplicação em contexto prático de defesa de sistemas. Sucedeu ao autor(*é frequente em contexto de trabalho*) e é um caso prático e que se considerou importante o suficiente para aqui constar. Foi um caso ***real***, ao serviço de uma entidade na qual trabalha/trabalhou (*propositadamente incerto*) e em que utilizando meramente OSINT, foi possível chegar à identidade do "potencial" agressor. Daí, utilizando também apenas OSINT, foi possível obter **TODOS OS DADOS** do atacante, incluindo nome, morada, estudos, histórico de vida e profissão actual.

Resumindo, foi detectado um ataque de injecção de SQL. Foi depois pedido à secção de *firewall*, a verificação do endereço IP atacante de origem brasileira[150]. O IP não era de VPN nem constava em nenhum site onde são informados/reportados abusos.

[150] Por experiência pessoal do autor, os portugueses e suas instituições são, desde há muito, escolhidos como alvo, quer de testes quer de ataques reais por parte de hackers de origem brasileira. Há a noção e a "falsa segurança" que os portugueses não dão muito valor à segurança informática, pelo que muitos atacantes nem se dão ao trabalho de usar VPN para encobrir os rastos.

Pegando unicamente em OSINT, foi pesquisado o IP, o que se podia saber a nível de DNS, entre outros, nomeadamente WHOIS, e de onde vinha (geoip - geolocalização).

Pegou-se também no ip e colocou-se no navegador web, tentando ver se ainda estava activo e se tinha alguma serviço web activado (e tinha!!). O que foi feito como investigação, se fosse com o objectivo de perseguir ou para conhecer alguém, era *cyberstalking*[151] (considerado crime).

Foi estudado o código *html* desse sítio web e concluiu-se que o site era criação do próprio atacante, portanto um site único. Verificou-se que o site apenas tinha "painel de controle" no nome, mas que usava "usuário" e password no formulário.

Continuando a análise do código da página *web*, verificou-se ainda que existia um titulo bastante peculiar. Foi procurado em motores de busca comuns e agregadores. Foram também procurados sitios de programação e partilha de código.

Foi localizado então um endereço web, ***.asuscomm.com. Seguida essa hiperligação, obteve-se um endereço igual ao sítio web que já tínhamos. Resultado: feito um *dnslookup* ao domínio, obtivemos o ip que já tínhamos (*BINGO*!). O nome do novo domínio encontrado, era *asuscomm*, ou seja, um domínio que a ASUS permite usar, aos utilizadores que usem os seus equipamentos NAS.

Baseado no nome *asuscomm*, é com muito forte probabilidade que aquele IP seja de um equipamento NAS, mais concretamente um ASUSTOR. (Existe um DNS da ASUS, que pode ser utilizado pelos NAS da própria ASUS, para que o dono do NAS e quem conhece e ter permissões para aceder ao NAS, não tenha de decorar um IP.)

O atacante tem o site alojado num NAS da ASUS, no sítio de onde lança o ataque? Não sendo muito inteligente, podemos ter aqui duas hipóteses:

a. o atacante ignora que alguém faça o que fiz e faça o *portscan* sem problemas
b. o atacante seja afinal vitima, e tenha sido o equipamento, parte de *botnet* ou comprometida, a fazer o ataque.

Vamos continuar...
Repetindo o M.O., voltou-se a pesquisar e obteve-se o nome *xxxxx*:

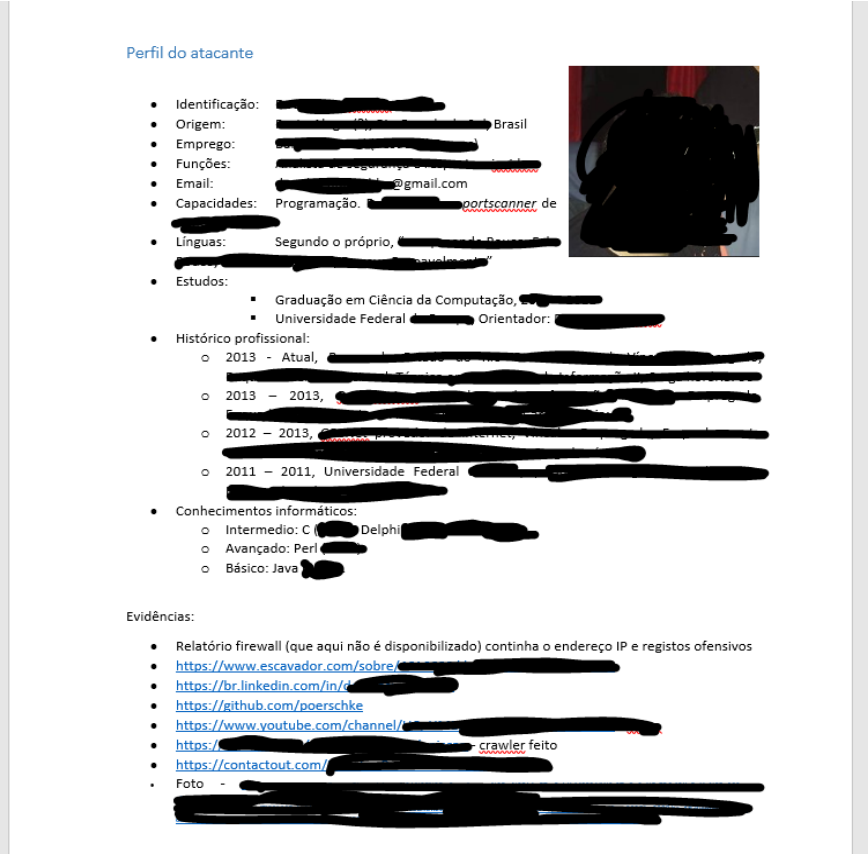
- em vários sites
- em 1 vídeo no *Youtube* (interessante o vídeo porque o "amigo" xxxx ensina a fazer um *portscan*…. Portanto confirma a culpa e o ataque)
- em sites de jogos. Por exemplo, é jogador de *xxx of xxxx*, entre outros
- em uma universidade, incluindo os seus orientadores de curso
- usa *linkedin*, onde tem foto, empregos, localizações, etc
- outros

Nesta altura sabemos (*Youtube*) que o *xxxx* criou um script de *portscan*, faz publicidade do mesmo no *youtube*, tirou um curso superior de informática, sabemos onde estudou, onde mora, quem foram os seus professores, etc.

[151] *Stalking* – expressão oriunda da palavra inglesa *stalk* que significa perseguir. *Cyberstalking* consiste na perseguição virtual, como por exemplo na alegação de falsas acusações, monitorização, ameaças, roubo de identidade, dano a dados ou equipamentos, solicitação de sexo, aquisição de informações para uso prejudicial.

No site do *SourceForge*, foi possível ver o projecto, assim como no *GitHub*. Vem com isso a identificação, o mail e outros projectos (este é mesmo o único).
Foi descoberto nome do atacante, o que programa, que tem conta na *sourceforge* onde partilha uma ferramenta que criou e que fez um *crawler* e usa o NAS xxx para recolher e armazenar essa informação.

Tendo descoberta toda esta informação, foi compilada de forma a traçar um perfil (brincadeira) do "potencial" agressor (última parte deste "*how-to*") e um relatório.



Perfil do atacante

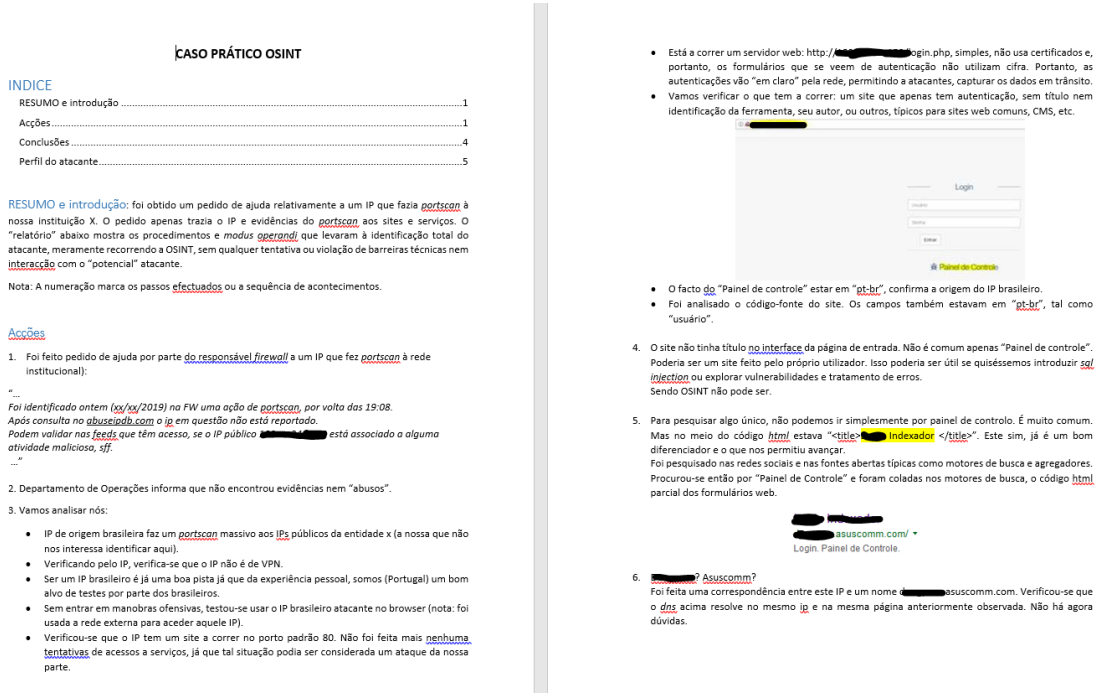
- Identificação:
- Origem: Brasil
- Emprego:
- Funções:
- Email: @gmail.com
- Capacidades: Programação. portscanner de
- Línguas: Segundo o próprio,
- Estudos:
  - Graduação em Ciência da Computação,
  - Universidade Federal Orientador:
- Histórico profissional:
  - 2013 - Atual,
  - 2013 – 2013,
  - 2012 – 2013,
  - 2011 – 2011, Universidade Federal
- Conhecimentos informáticos:
  - Intermedio: C (Delphi
  - Avançado: Perl
  - Básico: Java

Evidências:

- Relatório firewall (que aqui não é disponibilizado) continha o endereço IP e registos ofensivos
- https://www.escavador.com/sobre/
- https://br.linkedin.com/in/
- https://github.com/poerschke
- https://www.youtube.com/channel/
- https: crawler feito
- https://contactout.com/
- Foto -

Figura 16 - Construção do relatório e perfil do atacante

Em baixo, o relatório criado com o perfil do atacante. O relatório teve como objectivo ser um caso de estudo, sennsibilizando as pessoas para o OSINT e suas capacidades, assim como mostrar o perigo de partilha de informações pessoais na Internet. O relatório e evidências recolhidas poderiam ter sido entregues à policia para que estes fizessem o trabalho de investigação possível.



CASO PRÁTICO OSINT

INDICE

RESUMO e introdução: foi obtido um pedido de ajuda relativamente a um IP que fazia *portscan* à nossa instituição X. O pedido apenas trazia o IP e evidências do *portscan* aos sites e serviços. O "relatório" abaixo mostra os procedimentos e *modus operandi* que levaram à identificação total do atacante, meramente recorrendo a OSINT, sem qualquer tentativa ou violação de barreiras técnicas nem interacção com o "potencial" atacante.

Nota: A numeração marca os passos efectuados ou a sequência de acontecimentos.

Acções

1. Foi feito pedido de ajuda por parte do responsável *firewall* a um IP que fez *portscan* à rede institucional):

"...
Foi identificado ontem (xx/xx/2019) na FW uma ação de portscan, por volta das 19:08.
Após consulta no abuseipdb.com o ip em questão não está reportado.
Podem validar nas feeds que têm acesso, se o IP público [redacted] está associado a alguma atividade maliciosa, sff.
..."

2. Departamento de Operações informa que não encontrou evidências nem "abusos".

3. Vamos analisar nós:

- IP de origem brasileira faz um *portscan* massivo aos IPs públicos da entidade x (a nossa que não nos interessa identificar aqui).
- Verificando pelo IP, verifica-se que o IP não é de VPN.
- Ser um IP brasileiro é já uma boa pista já que da experiência pessoal, somos (Portugal) um bom alvo de testes por parte dos brasileiros.
- Sem entrar em manobras ofensivas, testou-se usar o IP brasileiro atacante no browser (nota: foi usada a rede externa para aceder aquele IP).
- Verificou-se que o IP tem um site a correr no porto padrão 80. Não foi feita mais nenhuma tentativas de acessos a serviços, já que tal situação podia ser considerada um ataque da nossa parte.

- Está a correr um servidor web: http://[redacted]ogin.php, simples, não usa certificados e, portanto, os formulários que se veem de autenticação não utilizam cifra. Portanto, as autenticações vão "em claro" pela rede, permitindo a atacantes, capturar os dados em trânsito.
- Vamos verificar o que tem a correr: um site que apenas tem autenticação, sem título nem identificação da ferramenta, seu autor, ou outros, típicos para sites web comuns, CMS, etc.

- O facto do "Painel de controle" estar em "pt-br", confirma a origem do IP brasileiro.
- Foi analisado o código-fonte do site. Os campos também estavam em "pt-br", tal como "usuário".

4. O site não tinha título no interface da página de entrada. Não é comum apenas "Painel de controle". Poderia ser um site feito pelo próprio utilizador. Isso poderia ser útil se quiséssemos introduzir *sql injection* ou explorar vulnerabilidades e tratamento de erros.
Sendo OSINT não pode ser.

5. Para pesquisar algo único, não podemos ir simplesmente por painel de controlo. É muito comum. Mas no meio do código *html* estava "<title>[redacted] Indexador </title>". Este sim, já é um bom diferenciador e o que nos permitiu avançar.
Foi pesquisado nas redes sociais e nas fontes abertas típicas como motores de busca e agregadores. Procurou-se então por "Painel de Controle" e foram coladas nos motores de busca, o código html parcial dos formulários web.

[redacted]
[redacted]asuscomm.com/
Login. Painel de Controle.

6. [redacted]? Asuscomm?
Foi feita uma correspondência entre este IP e um nome d[redacted]asuscomm.com. Verificou-se que o *dns* acima resolve no mesmo ip e na mesma página anteriormente observada. Não há agora dúvidas.

Figura 17 - caso real de investigação OSINT

4.3 Projecto: plataforma de OSINT, iKNOW

No decorrer deste curso de Mestrado, foi proposto pelo mestrando, a dissertação/projecto no âmbito das informações em fontes abertas, OSINT, com uma ferramenta que auxiliasse a obtenção de OSINT sem violar qualquer Lei, com objectivos bem definidos.

Foi estudado com maior profundidade o que existia em ferramentas *Open Source* neste domínio OSINT *(mas também ciberespionagem),* assim como ferramentas comerciais: umas muito genéricas, outras muito específicas. Foram pensadas e criadas várias ferramentas no decorrer do projecto, sempre com o foco na obtenção de informação em fontes abertas.

Dado o interesse por informações, pretendeu-se criar algo novo e diferente, que pudesse colocar em prática diferentes conhecimentos. Algo que pudesse obter informações de forma automática, legal, de forma distribuída, e que pudesse ao mesmo tempo ser modular e integrável em futuros projectos. Acima de tudo, interessava que fosse útil e pudesse ser usado de forma automática para a geração de informações (no formato de "relatórios") que pudessem ajudar a tomar decisões (o objectivo da OSINT). A ferramenta deveria poder dar ao utilizador, a oportunidade de inserir também novas fontes, imagens, o que pudesse ser útil... Uma ferramenta em que pudéssemos inserir o que procuramos, o sítio (onde começar), e esta, de forma autónoma, fizesse o varrimento, obtivesse tudo o que fosse informação de valor, seguindo novas hiperligações/caminhos e fosse guardando o que fosse pesquisando, gerando no final, um relatório com o encontrado (ex: site e código-fonte, hiperligações, emails, datas, autores, percentagem de termos encontrados, imagens, meta-dados, entre outros).

O iKNOW inicialmente pensado deu origem a vários projectos e ferramentas. Um *crawler*, uma plataforma web (ferramenta dedicada à criação, partilha e distribuição de

relatórios), captura de capas de jornais, informações de sites, obtenção de informações e metadados de imagens entre muitos outros, mas todos OSINT e de interesse. Depois com a evolução, uma ferramenta de recolha mais manual mas também mais adaptada e no fim, mais rápida, mais fácil de se inserir informações, de analisar, complementar e gerar relatórios.

4.3.1 Funcionamento e Evolução

O iKNOW pretendeu ser desde o início, uma ferramenta que pudesse ser utilizada por qualquer pessoa. Sofreu diversas mudanças conforme será a seguir visto, em função não só das necessidades de obter informação mas também do *feedback* dos utilizadores. Mais do que um *crawler*, ou que fosse esteticamente interessante ou fácil de usar, pretendia-se algo que realmente ajudasse a compilar informação. Algo útil.

4.3.1.1 iKNOW versão 0.1

Primeiro pensada e projectada para funcionar em linha de comandos/terminal (devido à simplicidade e portabilidade), foi feita uma aplicação que através do pedido (via terminal) do utilizador, procurava num determinado sítio *web*, por palavras-chave, hiperligações, números de cartões de crédito, entre outros. Oferecia depois um pequeno sumário que era exportado para um ficheiro *html*. O funcionamento era local, bastante simples (para o utilizador final que ignora o que está por trás pelo menos), mas muito versátil para quem usa a linha de comandos. Para funcionar remotamente era necessário ligar via SSH para o equipamento onde o iKNOW estava instalado e aí, correr o *script*. O resultado/*output* era exportado como já referido (se pretendido e configurado podia enviar por email, cifrado).
O funcionamento é/era feito através de vários ficheiros de código ou *scripts*, que corriam no terminal e mostravam o resultado nesse terminal ou enviavam para uma base de dados e um ficheiro. O aspecto gráfico era pobre, mas funcionava e os resultados eram os esperados.
Esta versão/sistema não era distribuída mas já usava uma base de dados local que permitia a distribuição de tarefas. Depois do feedback não se avançou para o multi-utilizador e distribuição de tarefas. Não permitia o envio de imagens para obtenção de metadados nem gerava relatórios de várias notícias de sítios web diferentes. Os scripts foram feitos em *Python*, com o objectivo da portabilidade e suporte da própria linguagem e bibliotecas ricas.

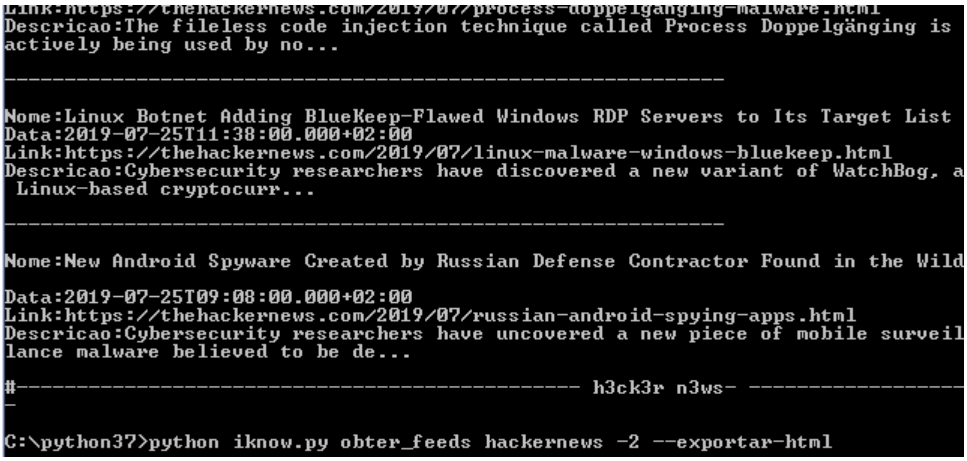


Figura 18 - iKNOW 0.1 – interface em linha de comandos (CLI)

4.3.1.2 iKNOW Versão 1.0

Revelou-se, no entanto, em testes de usabilidade(*ver inquéritos*[152]) que era difícil a "pessoas de fora", utilizar correctamente os comandos e compreender os resultados, tanto pela quantidade de informação criada, como pela dificuldade compreensível, de inserir os parâmetros que eram esperados, com as informações que se queriam obter, pouca legibilidade dada pelo terminal, e também pela impossibilidade de ter imagens e conteúdos web que fossem facilmente visíveis (*embora fossem exportados posteriormente com tudo*).

Interessa que além de útil, um utilizador comum, com quaisquer conhecimentos informáticos, não tenha muito trabalho a aprender a utilizar a ferramenta, não perca tempo, e não precise ter uma máquina Linux ou uma máquina dedicada. Daí se ter pensado em criar um servidor web, com os scripts todos pré-instalados e corridos de forma automática e no lado do servidor. (*A este respeito veja-se o capítulo de instalação e configuração, assim como o apêndice com o código-fonte em "7. Instação, código e configurações").*

Sendo agora uma plataforma web, foi adicionado um sistema multi-utilizador para que várias pessoas utilizem a ferramenta simultaneamente. Para que o utilizador se distinga de outros utilizadores e mantenha as suas próprias pesquisas separadas, foi criado um sistema de registo e autenticação.

Sendo simples, tentou-se algo ainda mais simples: utilizador entra no site e regista-se, utilizando um nome de utilizador único. O objectivo não é saber quem é o utilizador ou o que este pretende, mas sim, que cada utilizador tenha a sua própria informação sem interferir com a informação dos outros. O nome tem associado também um logotipo que aparece no relatório (*opcional do utilizador*).

A potencialidade de ser via web, abriu caminho à inclusão de imagens, tabelas e ficheiros. As operações que se pretendiam, podiam agora ser introduzidas no conforto do *browser* e não numa consola complicada.

Na imagem abaixo, figura 19, vemos a página de autenticação, onde o utilizador se pode registar e depois utilizar a plataforma.

[152] Inquéritos e seus resultados em "5.5 Testes, propostas e resultados dos inquéritos"

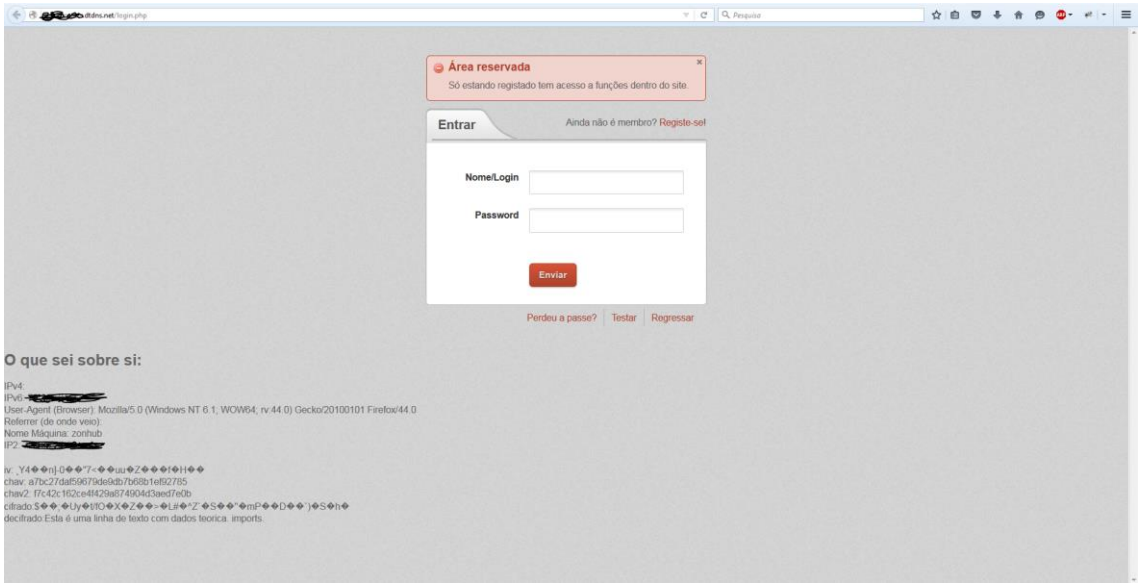


Figura 19 - iKnow 1.0 - Página de autenticação

Opcionalmente, pode fornecer uma imagem para se identificar (*figura 20*), e uma chave pública *GPG/PGP*[153] para receber no seu email, de forma cifrada e segura, notificações ou resultados de operações que entretanto sejam obtidas. A página pessoal é única para cada utilizador e mostra outras informações tais como o tempo que demora a carregar páginas, etc. Por razões de segurança, durante o funcionamento, apenas o próprio autor podia carregar páginas (*o autor consegue carregar páginas que se podem executar, não é um "bug" mas uma "feature"*).

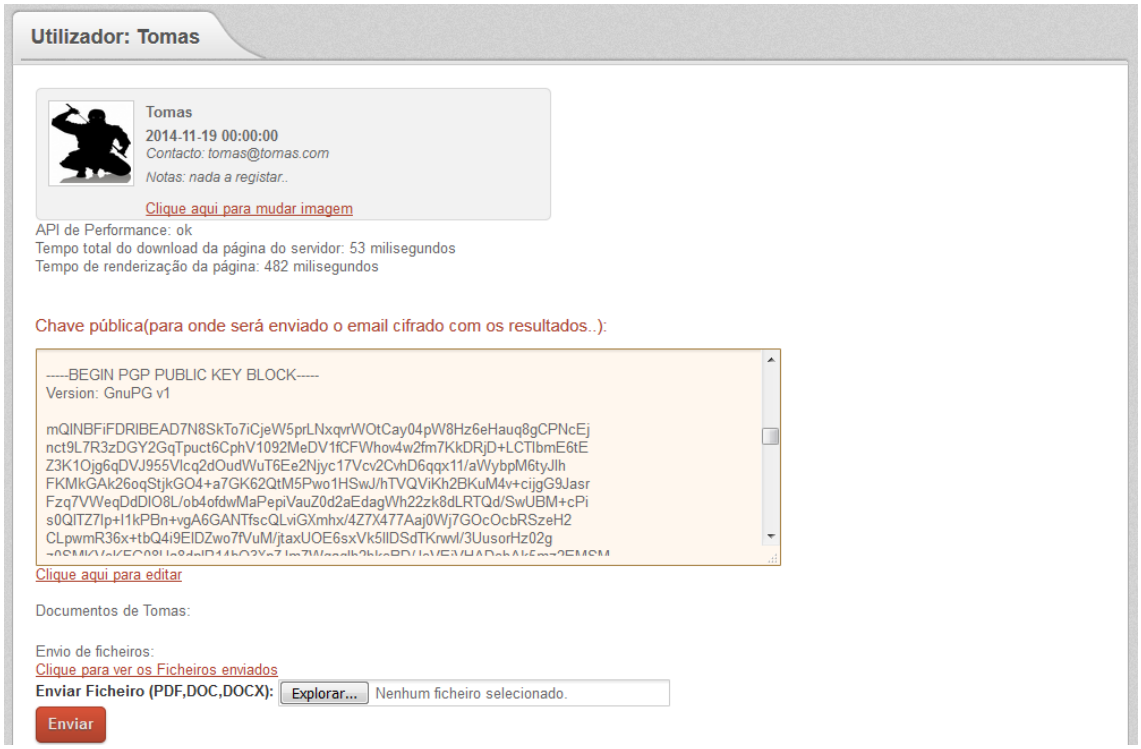


Figura 20 - Área Pessoal – Utilizador e chave pública RSA

[153] GPG - GNU Privacy Guard, *software* de encriptação compatível com o PGP (*Pretty Good Privacy*). Permite a geração e a troca de chaves assimétricas (diferentes para emissor e receptor)

Com esta versão, o utilizador é convidado a contribuir com poder de computação e localização para o projecto, podendo adicionar o seu próprio equipamento *raspberry* para todos os equipamentos funcionarem como um único/todo, distribuindo IP's, carga e localizações geográficas por todos, evitando assim que alguns dos equipamentos sejam bloqueados por fazer demasiados pedidos. O utilizador pode adicionar um equipamento e depois monitorizá-lo recorrendo à própria plataforma. Poderá ver de qualquer lugar o seu equipamento... Em baixo, do lado esquerdo as máquinas existentes (bem como o seu estado (activo ou *offline*), disponibilidade, temperatura, entre outros) no projecto e do lado direito, a disponibilidade online das mesmas (*mais informações sobre isto nos anexos).*

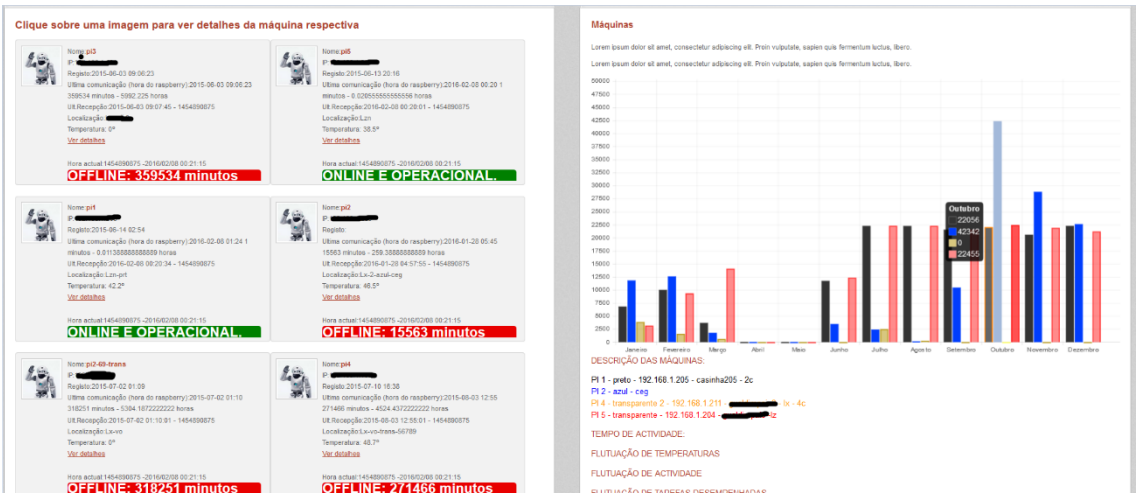


Figura 21 - Máquinas - Listagem de máquinas e sua Disponibilidade

Relativamente à navegação, todas as páginas do site contêm um menu superior que ajuda o visitante a não se perder durante a sua visita. Cada categoria tem as suas próprias mini-categorias, relacionadas com o tema principal.

Foram chamadas de "operações", as pesquisas efectuadas automaticamente pelo iKNOW, pedidas pelo utilizador. A operação (*ver figura 22*) é criada, e nessa mesma página é escolhido o endereço inicial do *crawler/scraper,* é determinada a quantidade de páginas máximas que podem ser obtidas, assim como se são apenas no próprio site/domínio, ou se pode "saltar" para domínios externos. Isto é importante definir porque se nada dissermos, podemos correr o risco do IKNOW pesquisar toda a Internet e nunca mais parar (teoricamente, pois pode ser desligado ou pode encravar ou o disco encher).

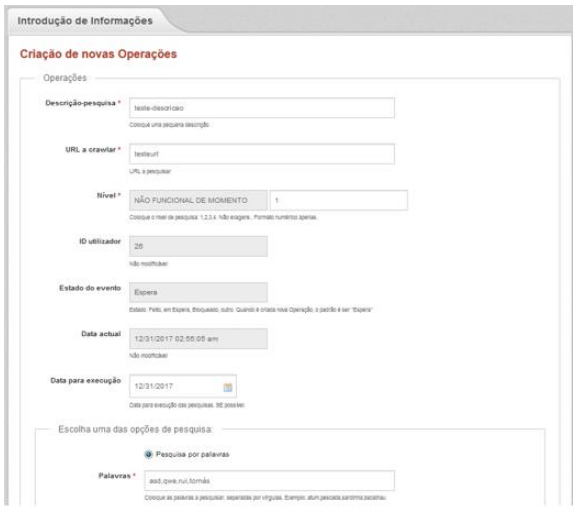


Figura 22 - Introdução de operações

O IKNOW vai começara a pesquisar tudo o que se pretende após poucos minutos. Quando houver resultados, a página vai sendo preenchida e o utilizador poderá consultá-la quando quiser. Também é possível ver as operações efectuadas no total ou por utilizador.

Figura 23 - Estatísticas globais e de utilizador

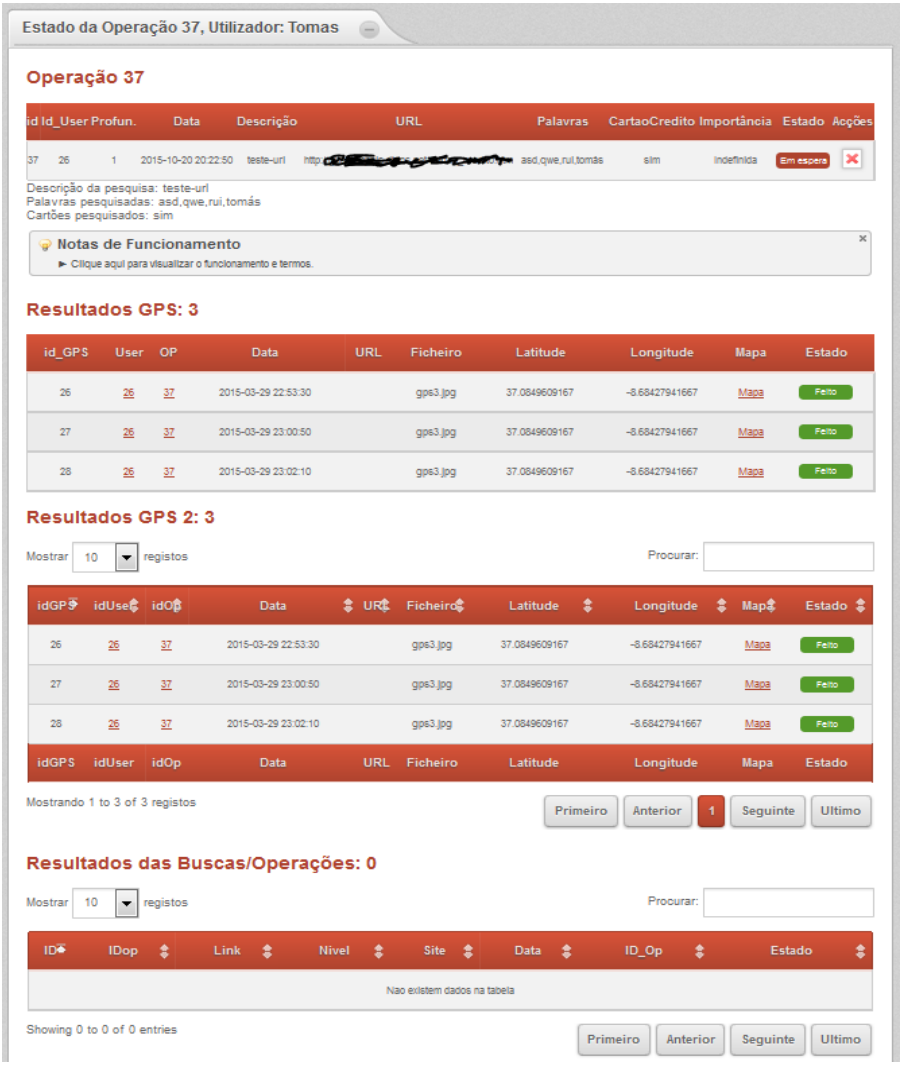


Figura 24 - Operações – Resultados obtidos

As operações vão conter as páginas encontradas (*apenas resultados positivo*), coordenadas GPS se existirem, além de um mapa do *Google Maps* com a geolocalização de alguma foto. Clicando em cima da hiperligação "Mapa", a aplicação abre-nos o navegador de internet no *site* do *GoogleMaps*, mostrando-nos o local.

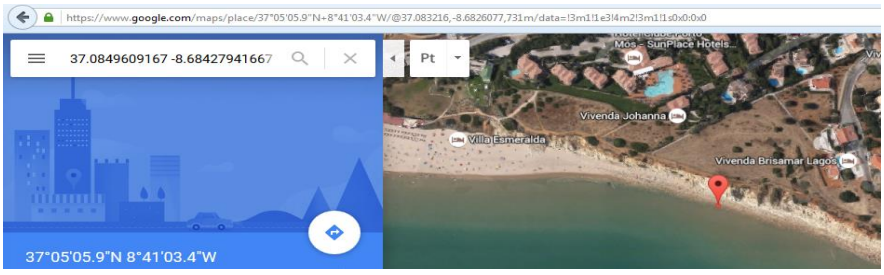


Figura 25 - Coordenadas obtidas de metadados de imagem são apresentadas no Google Maps

Apenas por razões académicas, foi criada uma área no sítio web IKNOW, que possui o código-fonte da aplicação, do servidor, e alguns extras, nomeadamente, código e imagens de como fazer com que o *Raspberry* acenda um led verde quando detectou palavras-chave e pisque

vermelho quando está em actividade, entre outros (temperaturas do próprio sistema, chaves assimétricas para comunicações em segurança, controlo remoto, etc).

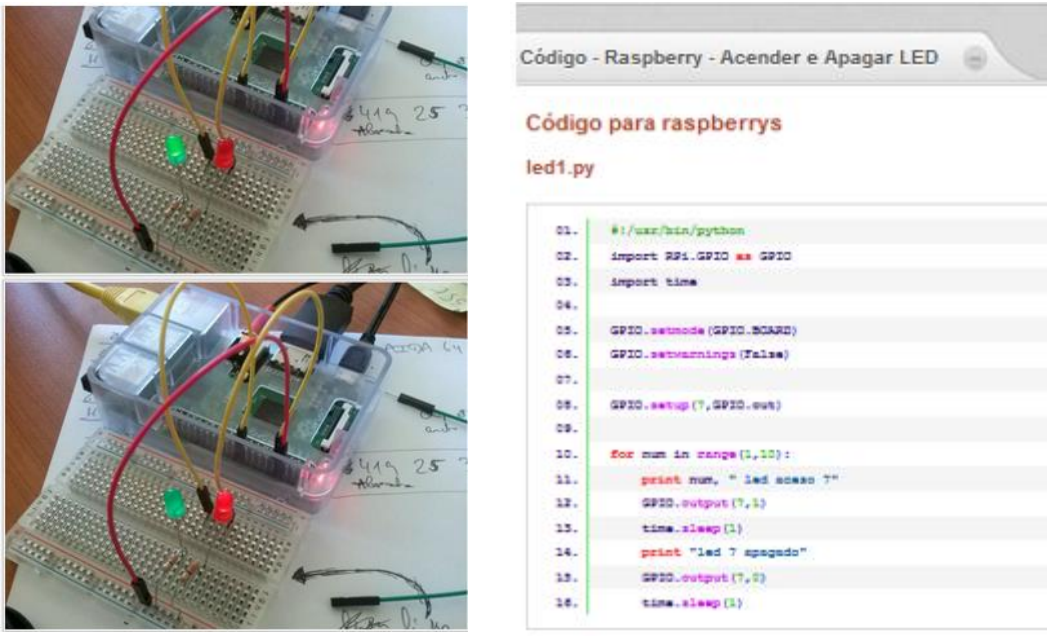


Figura 26 - Zona académica – Códigos variados para acrescentar funcionalidades. Aqui, código e imagem para acender LED's aquando houver descobertas

Foi criado de raiz um sistema de *tickets*[154]. A ideia foi/é, os utilizadores poderem criar tickets, distribui-los e resolvê-los entre si. O objectivo foi a entreajuda entre utilizadores e os seus Raspberrys/sensores. Também com este objectivo, foi, entretanto, alterado o funcionamento para que os utilizadores possam de alguma forma interagir e ver-se entre si.

Além da potencial utilidade, foi muito útil para o desenvolvimento do próprio site, actuando como uma "espécie" de GIT.

[154] Tickets – pedidos e/ou resolução de problemas

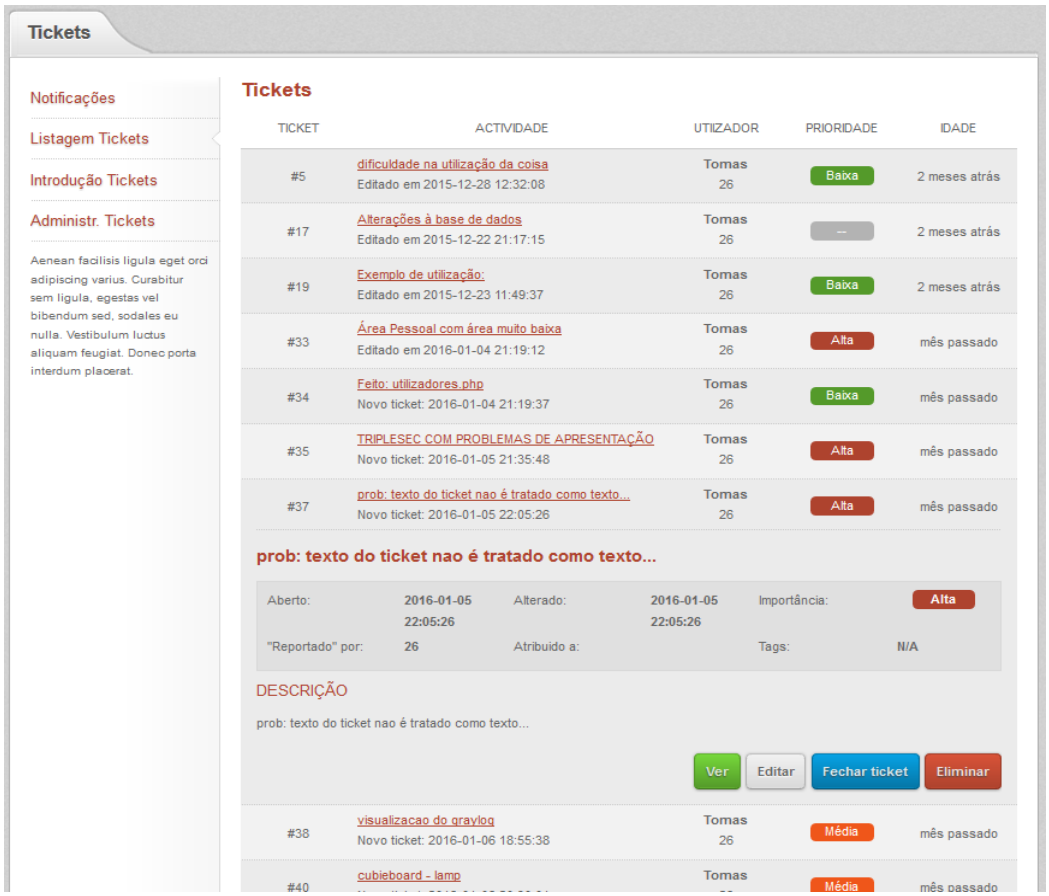


Figura 27 - Ferramenta de tickets iKNOW criada de raiz para o projecto

Foi criada uma secção de utilizadores, que nada mais é, que uma listagem de utilizadores (*figura 28).* Permite ver quem os utilizadores e os sistemas controlados (*opcional por utilizador, e inactivo por defeito, neste momento apenas o criador do site pode ver quem controla o quê*). Permite também ver os utilizadores bloqueados e o total de tarefas que têm (*não permite ver o quê*).

A secção dos utilizadores bloqueados foi criada devido a utilizadores-fantasma que apareceram e que tentaram fazer injecção de código e outros ataques ao sitio web, mesmo este não tendo publicidade nem estando publicamente disponível. A secção de *logs* no entanto identificou ips, browsers, e outras tantas informações que poderiam ser úteis *forensicamente*.

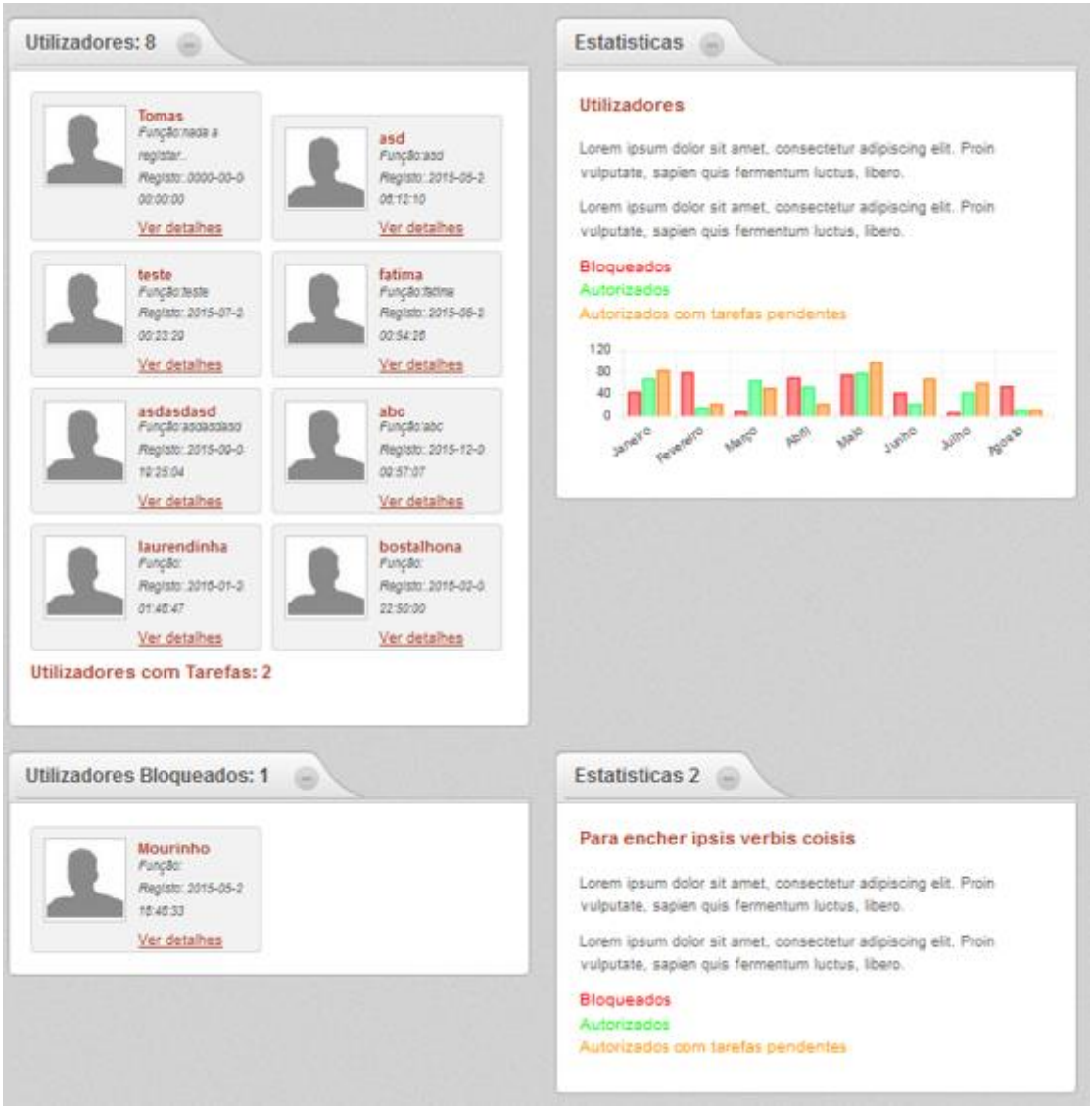


Figura 28 - Utilizadores - Listagem, Estatísticas e Bloqueios

Área de visitantes permite observar quem acedeu ao site, que navegador web utilizou (*user agent*). Apenas disponível para o administrador do site.

Figura 29 - Total de visitantes registados: 5476

A plataforma cresceu muito ao longo do tempo e contém muitas mais coisas. O foco no entanto são as operações e a obtenção de informação OSINT.

4.3.1.3 iKNOW Versão 2.0

Após diversos testes com utilizadores[155] (*incluindo testes que já se consideravam finais*), verificou-se que tão ou mais importante que o *crawler* ou que o funcionamento e monitorização dos Raspberrys com as suas estatísticas e controlo, ou que a geolocalização ou que as mil coisas, era um sistema que fizesse o que a versão 1.0 tinha como "*beta*": um formulário web que permitia baixar a noticia, directa e rapidamente. Também de extrema importância, e de alguma falta, era a capacidade/possibilidade de se inserir rapidamente e de forma cifrada, informações no site. Isto é agora possível recorrendo ao *bot*(automatismo) criado para o *software* gratuito e muito utilizado *Telegrama*[156].

Os utilizadores acharam útil baixar a noticia sem visitar o site e poder guardar essa informação. Se além de a guardar, esta pudesse ser armazenada e inserida num relatório, tanto melhor. Deu-se assim origem à versão 2.0 do iKNOW.

4.3.1.3.1 iKNOW 2.0 – Navegação sumária pela plataforma

Figura 30 - iKNOW 2.0 – Principais objectivos: obter e enviar

Novo interface. Todas as opções de relevo estão na primeira página. O objectivo é fazer tudo o que é importante a partir do mesmo sítio. Na imagem acima temos a capacidade de obter notícias de um site, e enviar capas de artigos, jornais, etc. O resto das opções são caminhos para novas páginas ou atalhos para geração, visualização de relatórios, entre outros.

[155] Ver inquéritos de utilização efectuados e seus resultados.

[156] Aplicação de mensagens e troca de ficheiros, que promete segurança, rapidez e anonimidade. Funciona online e está disponível gratuitamente no seu site em https://telegram.org/

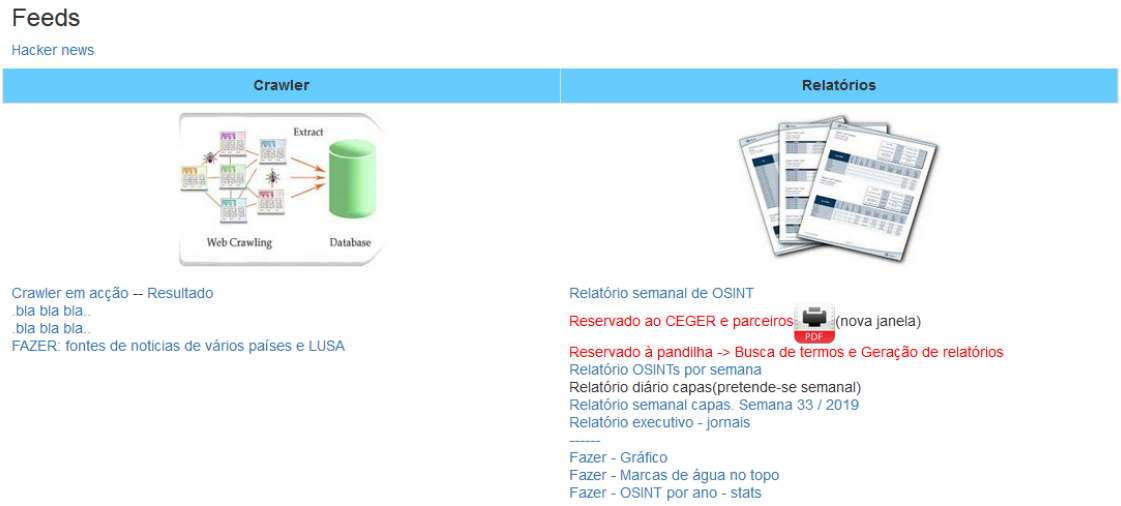


Figura 31 - iKNOW 2.0 - Feeds e relatórios

Na imagem acima temos a continuação da página anterior. Desta vez vemos as opções de: *crawler* PHP criado para este efeito, seu resultado, *feeds* de alguns sites como o *TheHackerNews*[157], criação de relatórios por semana, criação de relatórios por jornais entre outros como por exemplo, criar um relatório para um grupo de entidades (*em desenvolvimento*).

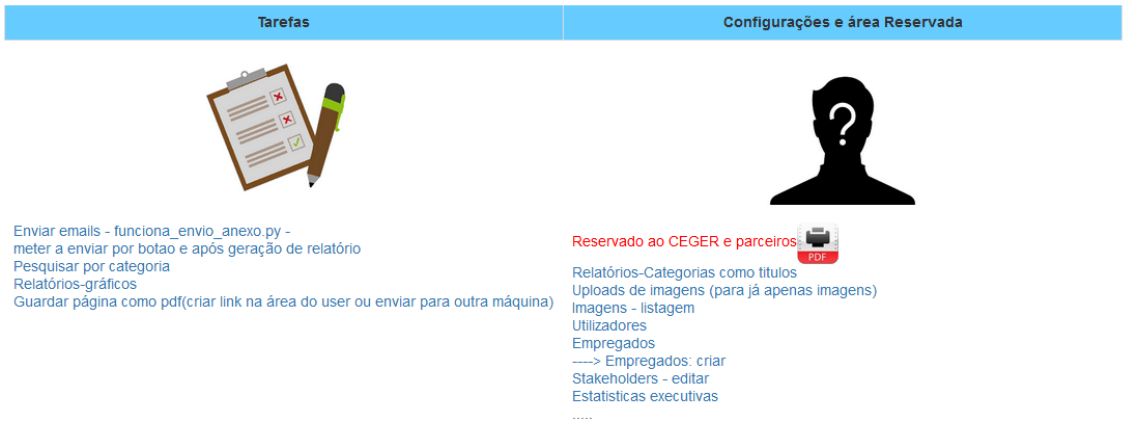


Figura 32- iKNOW 2.0 - Tarefas, Configurações e Área reservada

A categoria "tarefas" é uma zona de testes. Não está acessível ao utilizador comum. Configurações e área reservada é a área que permite a criação de categorias para relatórios, criação de entidades, envio de imagens para as entidades, entre outras coisas. Está em desenvolvimento para ser de facto multi-utilizador e permitir a partilha mais granular. Funciona.

[157] Sitio web de notícias relacionadas com segurança informática, disponível em https://thehackernews.com/

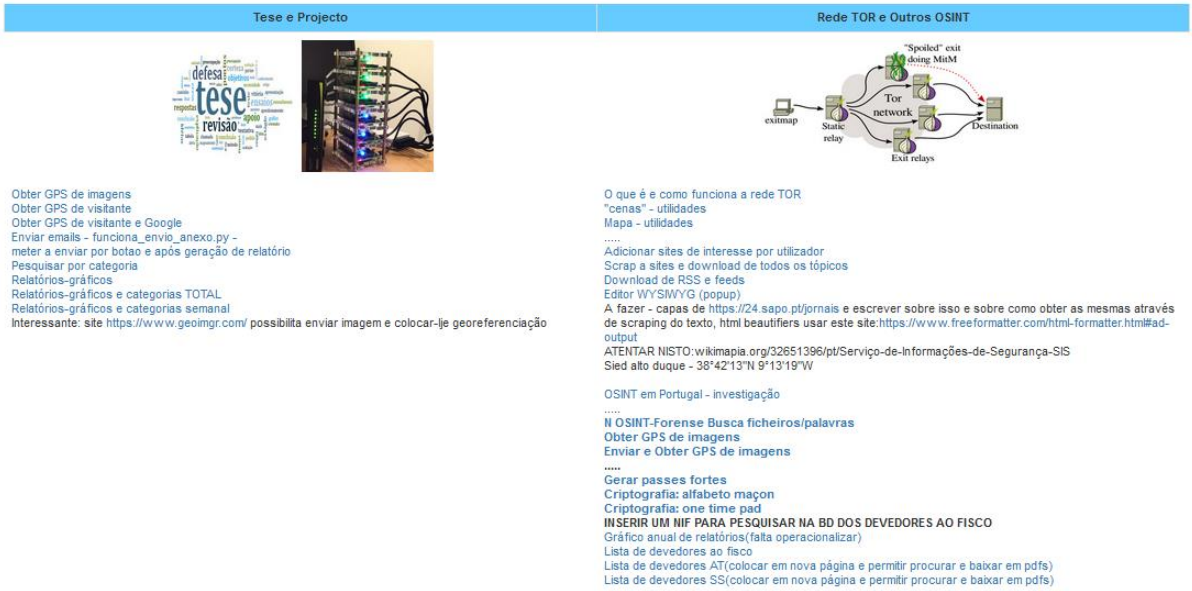


Figura 32 - iKNOW 2.0 - tese, tor e outros OSINT

Por fim, temos "Tese e Projecto", Rede TOR e outros OSINT. Tese e Projecto são um conjunto de mini-projectos para os quais se fizeram algumas mini-ferramentas que foram incluídas no site ou que foram eliminadas, mas que ficou a ideia. Duas delas que podem ser vistas na imagem acima, passam pela obtenção de meta-dados de GPS através de uma fotografia. A outra é através do *browser* e permissões do equipamento do visitante. Na versão iKNOW anterior o envio de notificações estava a funcionar muito bem porque partia do servidor. Agora pretende-se que seja o dono do equipamento/sensor/cliente/*Raspberry Pi*, a colocar os seus dados no seu equipamento.

Rede TOR e outros OSINT, são um conjunto de mini-ferramentas que podem impulsionar as capacidades do iKNOW. Para já, são usadas como prova de conceito, mas o objectivo é ser pedagógico.

Nota: algumas categorias têm subcategorias, que não serão descritas aqui.

4.3.1.3.2 iKNOW 2.0 – Diferenciação

O novo iKNOW mantém a obrigatoriedade de registo e autenticação tal como o iknow 1.0, pelo que passamos às maiores diferenças. O iKNOW 2.0:

1. Tem na página principal todas as opções, tentando com isto diminuir a curva de aprendizagem e oferecendo directamente as principais funções;
2. É mais sóbrio e rápido na aquisição de conteúdos (notícias, imagens, capas de jornais e relatórios, metadados de imagens, entre outros). Conta para isto um formulário directo para aquisição da página pretendida;
3. Totalmente orientado a relatórios, capas de jornais, informações;

4. Ganhou zona para criação de relatórios, extracção de informação em páginas, entre outros;
5. Permite o envio de imagens e a obtenção de metadados incluindo coordenadas GPS;
6. Permite o envio de imagens e a sua anexação a relatórios semanais;
7. Permite (em modo texto), obter conteúdos do OLX (útil para procurar equipamentos roubados por exemplo);
8. Permite a obtenção de imagens e conteúdos de sites como a Europol, que por defeito não permitem a obtenção de informação directamente;
9. Perdeu as zonas de *tickets* e de *Raspberrys (está prevista a reposição destas, que eram as caracteristicas mais interessantes tanto do ponto de vista do autor como dos inquiridos);*
10. Inclui um novo módulo de recepção de informações via *Telegrama*, que as insere automaticamente nas notícias/informações do mês. Está a ser criado agora um método para ser avaliado o texto e ser colocado automaticamente na categoria correspondente.

Nota: tanto o iKNOW 1 como o 2, são capazes de obter informações de alguns sites que obrigam o utilizar a registar-se ou a pagar para ver, ou no caso de já se ter atingido o limite de acessos a artigos de jornais online como no exemplo abaixo.

Figura 33 - Sem iKNOW, atingido o limite de visualizações, não temos informação

Com o iKNOW podemos obter a noticia do site mesmo que apareça a informação que atingimos o limite e nos redirecione para outra página.

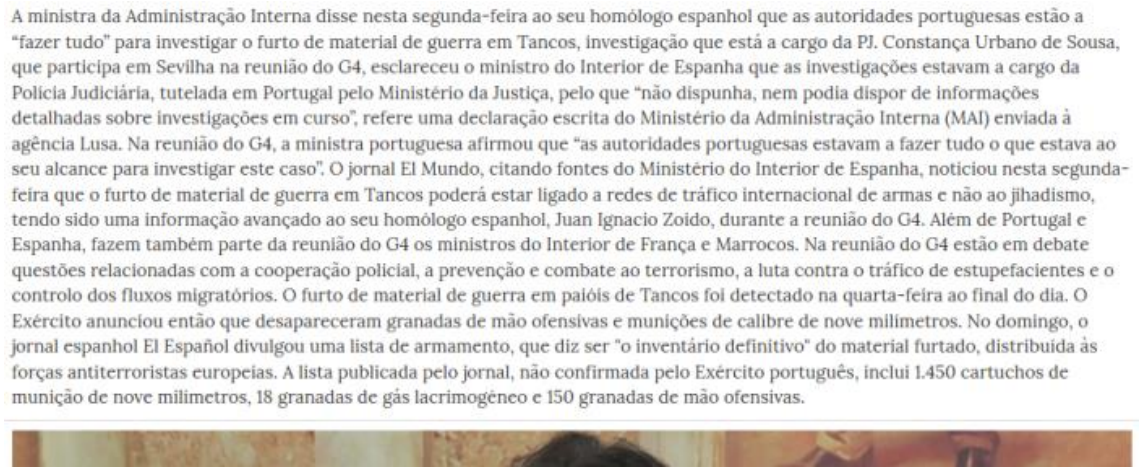

A ministra da Administração Interna disse nesta segunda-feira ao seu homólogo espanhol que as autoridades portuguesas estão a "fazer tudo" para investigar o furto de material de guerra em Tancos, investigação que está a cargo da PJ. Constança Urbano de Sousa, que participa em Sevilha na reunião do G4, esclareceu o ministro do Interior de Espanha que as investigações estavam a cargo da Polícia Judiciária, tutelada em Portugal pelo Ministério da Justiça, pelo que "não dispunha, nem podia dispor de informações detalhadas sobre investigações em curso", refere uma declaração escrita do Ministério da Administração Interna (MAI) enviada à agência Lusa. Na reunião do G4, a ministra portuguesa afirmou que "as autoridades portuguesas estavam a fazer tudo o que estava ao seu alcance para investigar este caso". O jornal El Mundo, citando fontes do Ministério do Interior de Espanha, noticiou nesta segunda-feira que o furto de material de guerra em Tancos poderá estar ligado a redes de tráfico internacional de armas e não ao jihadismo, tendo sido uma informação avançado ao seu homólogo espanhol, Juan Ignacio Zoido, durante a reunião do G4. Além de Portugal e Espanha, fazem também parte da reunião do G4 os ministros do Interior de França e Marrocos. Na reunião do G4 estão em debate questões relacionadas com a cooperação policial, a prevenção e combate ao terrorismo, a luta contra o tráfico de estupefacientes e o controlo dos fluxos migratórios. O furto de material de guerra em paióis de Tancos foi detectado na quarta-feira ao final do dia. O Exército anunciou então que desapareceram granadas de mão ofensivas e munições de calibre de nove milímetros. No domingo, o jornal espanhol El Español divulgou uma lista de armamento, que diz ser "o inventário definitivo" do material furtado, distribuída às forças antiterroristas europeias. A lista publicada pelo jornal, não confirmada pelo Exército português, inclui 1.450 cartuchos de munição de nove milímetros, 18 granadas de gás lacrimogéneo e 150 granadas de mão ofensivas.

Figura 34 - Com iKNOW, podemos ver o conteúdo mesmo que escondido

4.3.1.3.3 iKNOW 2.0 – Obter notícias de sítios web

"Obter noticia principal" é o principal destaque da página. Inserimos a URL pretendida e clicamos enviar. A menos que dê erro (*agora já tratados mas acontecendo, tal geralmente deve-se a caracteres não standard*), vamos ser enviados para uma nova página, auto-preenchida com tudo o que a página tinha, mas sem publicidade ou informação secundária.

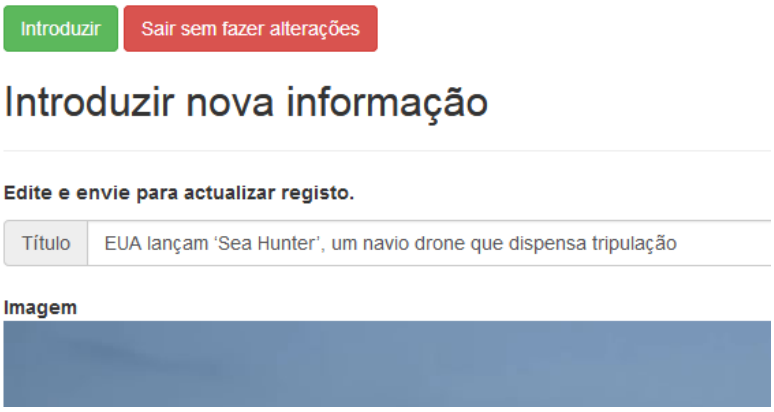
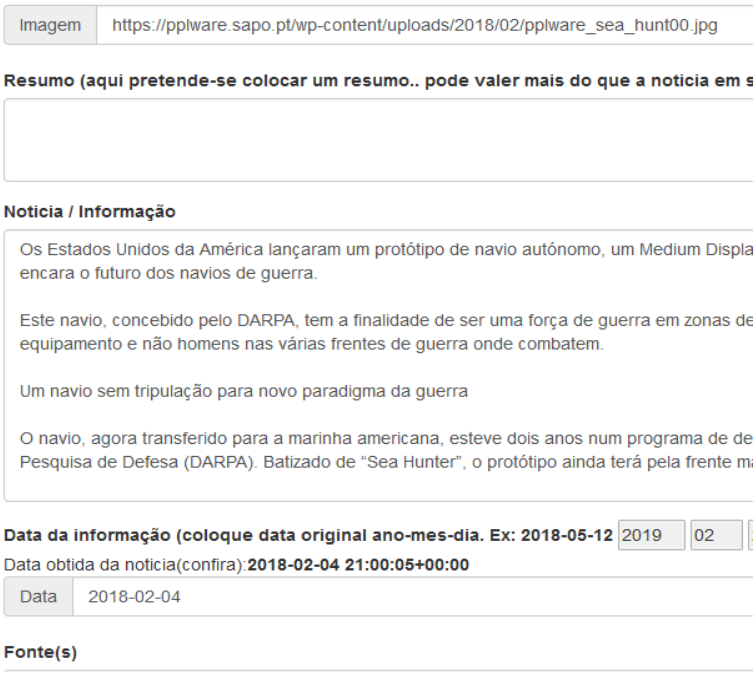


Figura 35 - Aceitar ou não? toda a informação é recolhida e o utilizador é que escolhe



O resumo está vazio pois essa informação deve ser o analista a preencher com a sua análise pessoal (e outros conhecimentos que tenha) da noticia.

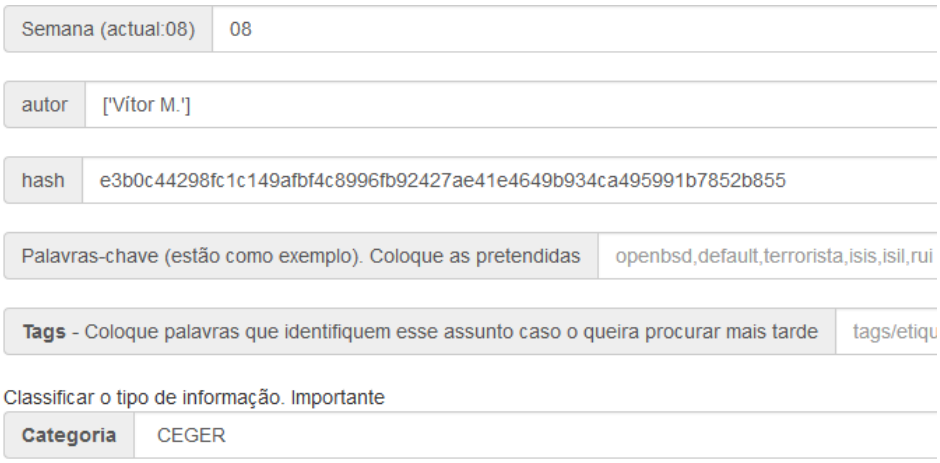


Figura 36 - Página mostra tudo o que foi recolhido. Pode ser alterado agora ou editado mais tarde

Se for aceite, toda a informação vai entrar para a totalidade de notícias, e mais interessante, no relatório da semana em que tivermos colocado a data (ou na sua ausência, na própria semana).

4.3.1.3.4 iKNOW 2.0 – Obter sítios da rede TOR

A obtenção de notícias via modo gráfico está restrita a sítios web ditos "normais". Na rede TOR, deve ser apenas o cliente a fazer esse pedido e a enviar essa informação para o servidor. Esse pedido é colocado na base de dados como estando para ser feito.

O primeiro *Raspberry Pi*/equipamento que perguntar à plataforma se "tem trabalho", recebe esse pedido. Executa e envia à plataforma a resposta (página web com tudo) para integrar a base de dados.

Se a página web contiver redireccionamentos ou muita informação, ou demorar muito tempo a responder (típico da rede TOR), existe um *timeout* e o pedido falha. Isto tem atrasado a implementação "a sério" pelo que ainda está a ser feito e testado. Não tem imagem por ser em modo de linha de comandos "sem muito para ver".

4.3.1.3.5 iKNOW 2.0 – Categorias e Resumos

A "Categoria" é inserida pelo utilizador, aquando da inserção da noticia ou posteriormente, de uma lista de categorias pré-existente. Se a categoria não for a adequada, pode adicionar uma nova, sendo que todos os utilizadores verão a nova categoria.

É importante a categoria pois é o grupo do relatório onde virá a informação. Também para efeitos estatísticos... O resumo geral, contém todas as informações e proveniência:

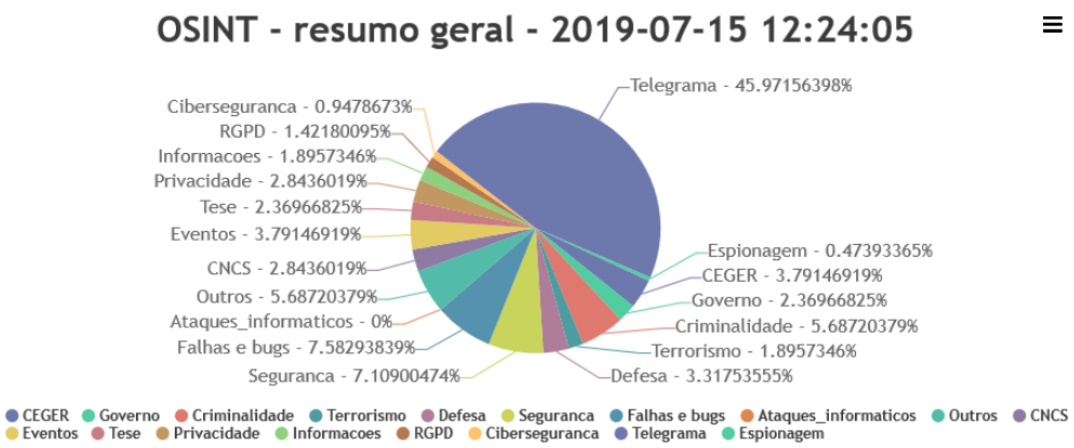
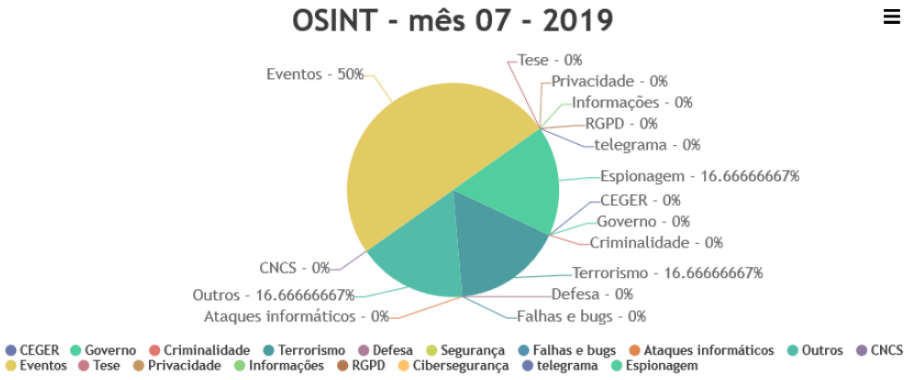


Figura 37 - Dashboards com os resultados OSINT obtidos, segundo quantidade e tema

No relatório mensal, são obtidas todas as informações relativas a um mês. No relatório semanal, apenas a própria semana, e é este relatório que nos interessa mais. Em baixo, o inicio do mês 7 de 2019, com poucas informações, mas com alguns eventos que decorreram ou vão decorrer.



4.3.1.3.6 iKNOW 2.0 – Obtenção e/ou criação de *feeds*

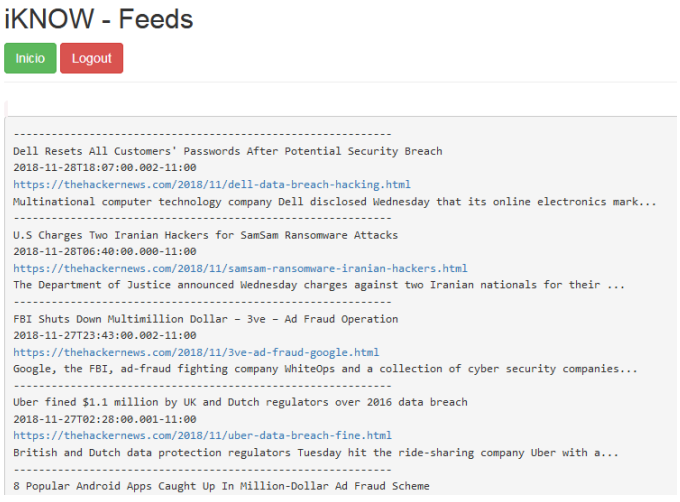


Figura 38 - feeds obtidos via scrap de sítios web

À semelhança dos IKNOW anteriores (1.0 e 0.1), também houve a necessidade que esta versão obtivesse de forma automática, *feeds* de *sites*, recolhidos pelo *crawler* e/ou por *scripts* criados do zero para este efeito. Estão a ser obtidos *feeds* de *sites* com *feeds*, mas também está a ser feito *parse* a sites que não fornecem estes dados, de forma a termos os nossos próprios *feeds*.

4.3.1.3.7 iKNOW 2.0 – Construção de relatórios OSINT

Uma das principais motivações, foi a construção de uma ferramenta que auxiliasse nos relatórios, procurando, mostrando, coleccionando e depois reunindo num documento(imagem 39 mostra o relatório na secção "Governo"). Para tal, a um dia da semana, é gerado um relatório que é composto pela soma de todos os documentos obtidos durante a semana. Documentos obtidos manual ou automaticamente. Neste momento é tudo manual devido à carga que é colocada.

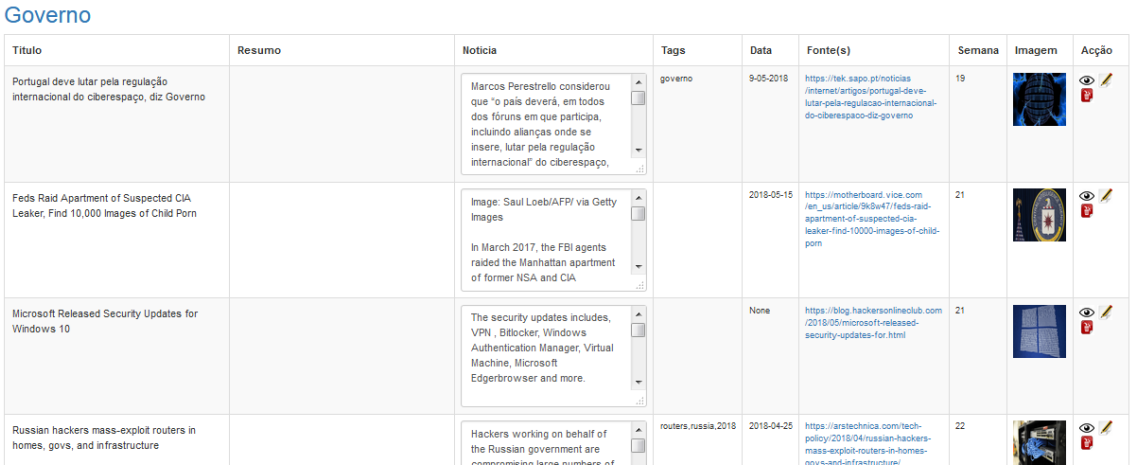


Figura 39 - iKNOW 2.0 - Construção de relatórios

Mesmo manualmente, a quantidade de informação chega a ser enorme. Para isso, muito tem contribuído a criação do bot de *Telegram* (ver *bot telegram*[158] mais à frente e sua construção[159] nos Apêndices). Depois de introduzida a informação, é criada uma listagem como mostra a imagem 39, dividida por categorias. Cabe ao utilizador seleccionar se quer tudo ou não, e se a categoria está correcta.

[158] Apêndice 7.5 Código do bot Telegram
[159] Apêndice 4.5.1 Recolha de informação via Telegrama 108

4.3.1.3.8 iKNOW 2.0 – Obtenção de capas de jornais

As capas de jornais são obtidas manualmente ou automaticamente. Neste momento há um *script* automático em CRON, que obtém automaticamente todos os dias às 9:30 as capas de jornais de alguns sítios web. Nem sempre funciona porque os sítios web estão constantemente a mudar as protecções (mudam nome do jornal ou colocar código numéricos para não se conseguir baixar a imagem, ou outros…).
As capas são de imensa utilidade para se ter uma ideia geral do estado do País, nomeadamente quanto à questão ciber. A figura 40 mostra o aspecto das capas de jornais obtidas por dia, pela ferramenta iKNOW.

Figura 40 - iKNOW 2.0 - Obtenção de capas de jornais

4.4 Requisitos para instalação da plataforma iKNOW

Como tudo na informática, são necessários alguns requisitos para poder instalar quer o servidor quer os clientes. Os requisitos necessários são todos eles de código-aberto (*open source*), gratuitos, e fáceis de obter. As necessidades diferem do cliente para o servidor, como é natural. Pretendeu-se desde o início, e como prova de conceito, que o projecto pudesse ser implementado por qualquer pessoa, que fosse barato e simples de perceber. Apenas se refere aqui os requisitos para os iKNOW 1 e 2.

4.4.1 Servidor

O servidor é o local físico e lógico onde estão armazenadas (*numa base de dados*) todas as informações obtidas, utilizadores, configurações técnicas, operações, interface gráfico, etc. É, portanto, mais exigente em termos de recursos computacionais e de largura de banda do que um cliente. No geral depende muito de quantos utilizadores vão aceder simultaneamente à plataforma e do número de registos que esta vai ter na base de dados.

Tudo o que o servidor e cliente(s) precisam, pode ser instalado via *script* ou seguindo os passos fornecidos. Precisamos no geral, de uma máquina com acesso físico e/ou SSH, e:

- Memória RAM suficiente[160] para poder servir páginas e fazer pesquisas na base de dados em simultâneo;
- Disco rígido com espaço suficiente[161] para armazenar informação (dados na forma de texto, imagens, ficheiros, …). Quanto mais rápido o disco, melhor;
- Uma plataforma iKNOW servidora precisa de:
    - Servidor web *NginX* (preferível ao Apache por ser mais leve);
    - Linguagem de *scripting PHP* para fazer ligação servidor web/base de dados;
    - Base de dados *MySQL*/*MariaDB* (armazenamento de informações) e *cluster Galera* (sincronização entre bases de dados *MySQL/MariaDB*);
    - Páginas web do iKNOW, scr*ipts web, scripts Python, scripts Bash*, … (tudo isto está nos anexos e página *GitHub*)
    - Um servidor Linux precisaria ter um conjunto LEMP (*Linux Nginx, MySQL*, *PHP*), facilmente instalável.
    - Um servidor Windows pode simplesmente baixar[162] este *software WAMP* (*Windows Apache*/*Nginx MySQL PHP*) de uma vez.
- Internet e largura de banda suficientes para servir q quantidade de utilizadores pretendidos. Quanto mais, melhor;
- Capacidade de processamento para executar as ordens, *scripts*, fazer sincronização de dados, processamento e acessos à base de dados, entre outros. Uma máquina minimamente recente é capaz desta tarefa;
- Interpretador de *Python*; Foi utilizado no servidor apenas *Python* versão 3.x.

4.4.2 Clientes

Os clientes são utilizados para obter os pedidos/operações do servidor e de seguida, colocá-los em acção, enviando depois ao servidor, todos os resultados obtidos. Nenhum cliente guarda informações do que for obtido após finalizar as suas tarefas. Os clientes são simultaneamente clientes (pedem informação), sensores (envio de informações de ambiente, estado de operação, temperaturas, entre outros) e agentes (respondem a ordens do servidor directamente e outros clientes indirectamente). É neles que tudo acontece.

[160] Esta quantidade varia em função do sistema operativo, servidor web, se utiliza páginas dinâmicas ou não. Para o iKNOW, foi testado em ssistemas raspberry e percebeu-se que1 giga é suficiente mas 2 ou 4 são melhores.

[161] Espaço em disco depende da quantidade de informação que se pretende vir a ter. Uma pessoa sozinha pode ter necessidade de apenas 100megas, mas vários utilizadores precisarem de muito mais

[162] *Download* gratuito em https://sourceforge.net/projects/xampp/files/XAMPP%20Windows/

O cliente está originalmente pensado para operar em *Raspberrys Pi.* Baratos, económicos em electricidade (e preço de aquisição), rápidos, a funcionar em Unix/Linux e com todo o suporte que precisamos para comunicar com a plataforma. Precisamos de:

- Uma máquina *Raspberry Pi 2, 3 ou 4* (mas também pode ser uma qualquer distribuição Linux) a actuar como agente ou sensor;
- *Python* versão 3.x (qualquer uma da versão 3);
- Internet;
- *Script* de instalação do iKNOW;
- Sistema operativo *Raspbian* (qualquer Unix/Linux serve mas os *scripts* têm por base este sistema (testado na distribuição Linux *Ubuntu*[163] e funcionou sem problemas));

4.5 Funcionamento

4.5.1 Recolha de informação via Telegrama

A aplicação Telegrama funciona em modo de "*app*" para telemóvel e em modo de aplicação para computador. Esta flexibilidade, juntamente com as suas promessas de comunicações cifradas fazem desta ferramenta a opção quase ideal para troca de mensagens entre utilizadores, mas também entre grupos (ou também chamados internamente de "canais").

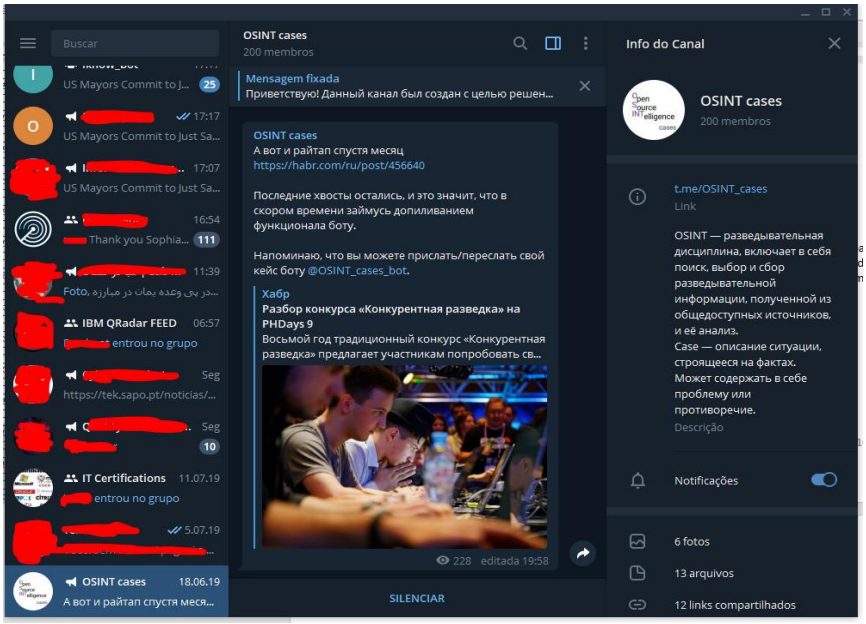


Figura 41 - Telegrama - um canal de conversação/chat

Na imagem 41, pode ver-se um canal típico de conversação. Na lateral esquerda os canais e pessoas, na lateral direita a identificação do canal, ao centro, a conversação propriamente dita. Foi criado (legalmente) para o iKNOW, um *bot*, uma aplicação que está permanentemente ligada e a escutar. Se alguém lhe enviar uma mensagem com uma hiperligação, ele guarda essa ligação, baixa a página e envia isso para o nosso servidor iKNOW, onde fica armazenado em base de dados, na categoria Telegrama, para ser revisto posteriormente. Na figura 42, abaixo, o *bot*,

[163] Ubuntu – distribuição de sistema operativo Linux. Mais informações online em https://ubuntu.com/

sempre a ouvir, obtém uma notícia e uma hiperligação que ele vai seguir, baixar e introduzir no iKNOW.

```
-------- Inserção e obtenção da informação na BD --------

d20c4632bec5dd44df006016404038c6a8caefe2d38ba55b9e52f30065122407
Data:2019-7-16
Autor(es):[]
1 Registo inserido. Posição  457


-------- Obtenção da informação via telegrama --------

        Link(s) : https://www.darkreading.com/attacks-breaches/us-mayors-commit-to-just-saying-no-to-ransomware/d/
        Data    : 16 07 2019 semana:29 hora:17:17:57
        Noticia : US Mayors Commit to Just Saying No to Ransomware
https://www.darkreading.com/attacks-breaches/us-mayors-commit-to-just-saying-no-to-ransomware/d/d-id/1335255?_mc=r

The group of more than 1,400 top elected municipal officials takes the admirable, recommended stance against payin
        sha256 da noticia:  5f2ccaeaf01ace91d69964b1d2ca9960837132ea4788ae00a6db36b6427d5330
erro ao conectar bd: 1062 (23000): Duplicate entry '5f2ccaeaf01ace91d69964b1d2ca9960837132ea4788ae00a6db36b6427d53
        Url:    https://www.darkreading.com/attacks-breaches/us-mayors-commit-to-just-saying-no-to-ransomware/d/d-


-------- Inserção e obtenção da informação na BD --------

cd9bee2d6cc10b1d54f31d59b56819cf8fdc58810391793aec2b344ee9abcb17
Data:2019-7-16
Autor(es):[]
1 Registo inserido. Posição  458
```

Figura 42 - O bot a funcionar do lado servidor

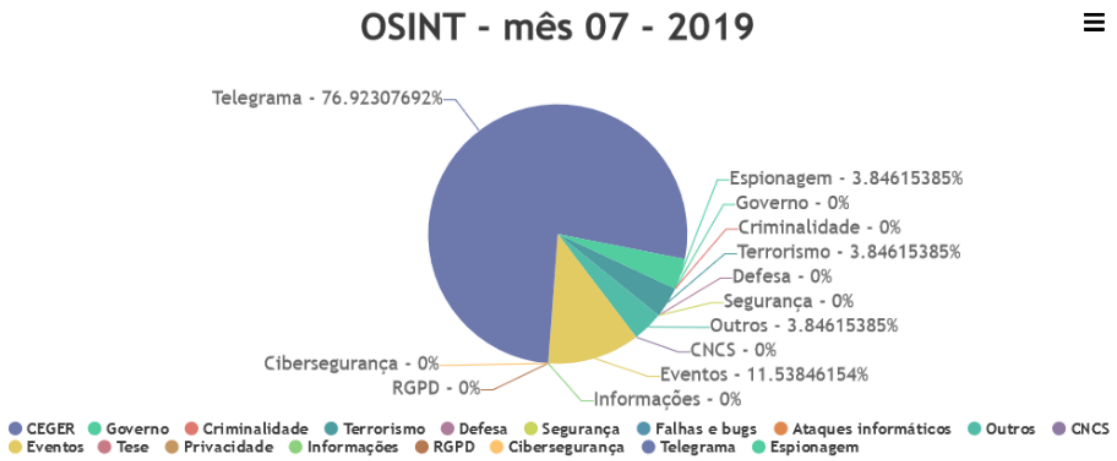


Figura 43 - iKNOW – informações do mês 7. Verifica-se a grande obtenção de notícias via Telegrama

A grande vantagem do *bot* criado para o iknow, usando o telegrama, é a privacidade e a rapidez com que rapidamente podemos inserir informações, seja com o telemóvel ou o PC, sem o utilizador estar preocupado com a forma como vai aceder ao site. Não precisa aceder.
Este automatismo(*bot*) veio poupar muito tempo e foi das características mais úteis, também solicitadas pelas pessoas que testaram o iKNOW. A sugestão inicial foi o *WhatsApp* mas o *Telegrama* veio a revelar-se uma aposta melhor, visto que os grupos e a anonimidade são melhores para privacidade e obtenção e facilidade de partilha de informações, além que a API do *Telegram* pode ser utilizada sem restrições e para diversos fins.

A importância deste bot é enorme já que é a maior fonte de informação a alimentar o iKNOW.
As notícias recebidas via Telegrama, chegam com este título e devem ser alteradas posteriormente para que a categoria seja diferente (*está a ser trabalhada uma alternativa para que no fim do texto conste a categoria que o iKNOW deve utilizar*).

Funcionamento:

O utilizador obtém uma hiperligação, uma notícia, algo que tenha um link. Partilha como faria normalmente com um telemóvel ou computador, escolhendo telegrama. Depois, escolhe o utilizador de telegrama "iknow_bot". Automaticamente o nosso *bot* recebe a informação, acede ao endereço, baixa a notícia, e insere na base de dados e no relatório semanal.

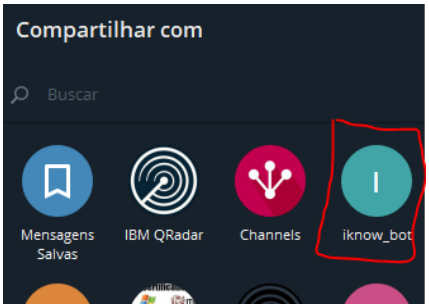
Em alternativa, o utilizador pode enviar uma noticia para o site enviando ao bot, a notícia, a categoria que deve ser inserida, a data, e quem envia.

O código fonte do *bot* foi escrita de forma simples, mas funcional e de acordo com os objectivos. Encontra-se nos apêndices ("8.5Código do *bot Telegram"*) o código-fonte.

Resumo gráfico:

1. Noticia tem interesse

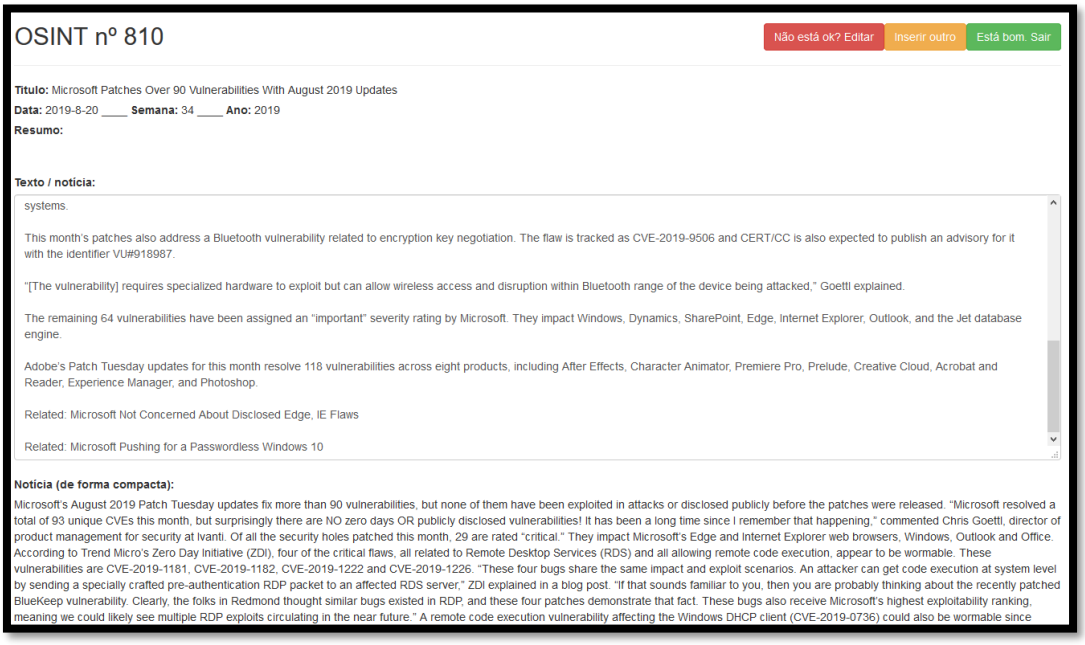
2. Partilha com o *bot*



3. App Telegram envia a informação para o bot, que por sua vez insere no servidor a informação que foi partilhada. Devolve "sucesso" (ou não) e a posição na base de dados.

```
-------- Inserção e obtenção da informação na BD --------
Titulo:Microsoft Patches Over 90 Vulnerabilities With August 2019 Updates
Data:2019-8-20
Autor(es):[]
Link:http://feedproxy.google.com/~r/Securityweek/~3/kb_kdEgRW0I/microsoft-patches-over-90-vulnerabilities-august-2019-u
dates
SHA256:  b7834425155a253a9faf6fe08871d7865219a439228aaed0ead8ebd0a71f6bb0
Hash do codigo_base64:b7834425155a253a9faf6fe08871d7865219a439228aaed0ead8ebd0a71f6bb0
1 Registo inserido. Posição  810
```

4. A noticia está agora no iKNOW e pode ser vista. Classificada na categoria certa e introduzida no relatório.



4.5.2 Recolha e processamento via *crawler* Python

A recolha de informações é a segunda[164] fase da metodologia OSINT. Deve ter em conta o que se pretende recolher e ter as ferramentas correctas para este efeito. Para recolha de informações em sítios web, é preciso ter em mente que cada sítio web é diferente, especialmente se não respeitar as correctas formatações *html*[165] *(tags,* títulos*,* códigos *css* e afins) ou se não tiver sido construído de forma coerente. Este é um desafio que irá estar sempre presente, assim como a vontade cada vez maior de dificultar aos *crawlers* a obtenção de informação (por parte dos donos dos conteúdos/sítios web).
A ideia inicial da ferramenta iKNOW funcionar assente sobre um *crawler* foi devido à necessidade de automatização da ferramenta. Pretendia-se que esta pudesse saltar da página web inicial, para outros sítios web e por aí adiante, baseada em hiperligações, ficheiros ou hipertexto. A mera ideia de criar um *crawler* foi um desafio muito interessante. As potencialidades de se varrer a internet e apanhar informação de forma automática, criar métricas, procurar termos, números e palavras-chave enquanto o fazemos, são de um potencial incrível. Pretendia-se "*crawlar*" todos os sites, a partir do site-alvo, com objectivos bem definidos, tais como a localização de palavras-chave, cartões de crédito, nomes de pessoas desaparecidas, entre outros.

Funcionamento: Obtenção das hiperligações visíveis e invisíveis (presentes no código, mas não possíveis de ser clicadas) na página inicial. O *crawler* percorre depois de forma ordenada e por "níveis de profundidade" todas as hiperligações dentro do próprio sítio web. As hiperligações para sites externos não são percorridas embora fiquem registadas na base de dados, podendo ser, a pedido do utilizador, mais tarde percorridas.

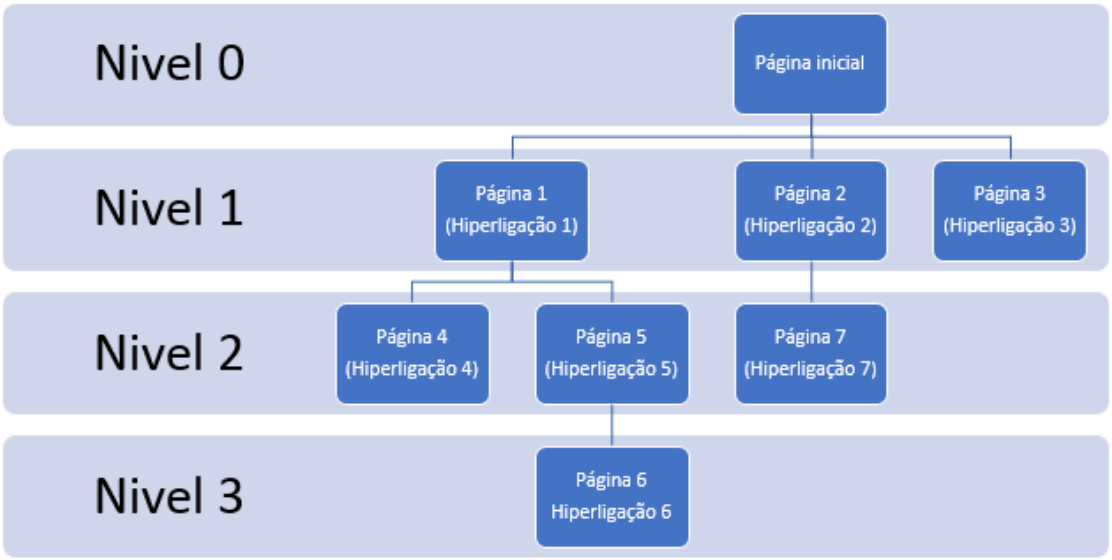


Figura 44 - Gráfico de funcionamento do crawler Python

[164] Sendo a primeira, o planeamento.
[165] HTML – *hypertext markup language*. Linguagem de programação que permite que texto, imagens, e dados sejam transformados e mostrados num navegador web, permitindo a navegação entre páginas, a utilização de diversos tipos de texto e a capacidade de pedido-resposta para servidores web.

A imagem 44, pretende exemplificar o comportamento referido. A opção desta técnica foi devido à simplicidade e objectivo de não percorrer sites fora do mesmo dominio.

Objectivo: tendo em conta as limitações e sendo este um projecto académico optou-se por criar o nosso próprio *crawler*, aprendendo e podendo também partilhar o realizado. O mais simples, mas funcional possível, sem preocupações por ora, com a velocidade e/ou eficiência. Antes de nos iniciarmos nesta aventura, estudámos outras soluções existentes, como por exemplo o muito conhecido *Scrapy*[166]. Simples, mas demasiado volumoso e de instalação difícil para o utilizador que queremos.

Foram criados dois *crawlers*, um em *PHP* e outro em *Python*. O funcionamento tinha por base um caminho inicial (url base do site pretendido), que indexa toda a informação contida no site, e que depois percorre todas as hiperligações que encontrar (assim como ficheiros pdf, imagens, tudo o que tenha hiperligações).

O indexamento do site é sempre feito, mas o armazenamento de informação apenas é feito se encontrar "x" percentagem dos termos procurados. Encontrando informação deve continuar. Não encontrando, deve procurar apenas mais "*x*" ligações e depois parar ao fim de determinada profundidade ou número de ligações (*caso o crawler não parasse, o armazenamento mais cedo ou mais tarde iria encher e deixar de funcionar*).

4.5.3 Recolha e processamento via *crawler* PHP

Foi utilizada uma única página web no iKNOW, com um exemplo de um *crawler* PHP[167]. Consta apenas por uma questão de prova de conceito. Não é, no entanto, a melhor solução para um projecto como o iKNOW. O PHP é uma linguagem fantástica e foi utilizada como ponte entre a interface e a base de dados. Tem uma curva de aprendizagem pequena, é fácil de utilizar, tem uma comunidade muito activa, entre outros.

Limitações: bastantes. *Timeouts*: no iKNOW teve de ser aumentado o tempo limite de resposta ou o utilizador teria os chamados *timeouts*, devido à aparente inactividade do sítio web, quando na prática o servidor ainda estava a reunir toda a informação antes de a apresentar no site (grandes volumes de informação retiradas da base de dados como na página de "resumo geral"); Outra limitação encontrada está na natureza evolutiva do PHP (quando se iniciou o projecto estávamos na versão 5 e agora na versão 7, o código do *crawler* teve de ser todo alterado e perdeu funcionalidades, inclusive bibliotecas que ficaram *deprecated*); Ainda outra limitação prática prende-se com o objectivo do projecto: queremos que o trabalho de *crawling*, obtenção de dados e processamento seja distribuído e não que seja o servidor a fazê-lo, correndo o risco de ser bloqueado pelo(s) site(s) que está a obter informação.

Funcionamento: resumidamente, o *crawler* PHP está inserido numa página típica PHP que verifica se utilizador está ou não autenticado. Estando, pega na página/caminho URL que lhe foi

[166] https://scrapy.org/

[167] PHP – acrónimo para *Hypertext Preprocessor.* Linguagem de programação dinâmica que actualmente é utilizada em milhões de sítios web. É utilizada no iKNOW para mostrar gráficos, manter sessões, fazer a autenticação e a ponte entre a base de dados e a interface gráfica.

dado e segue essa página, baixando todo o conteúdo e depois abrindo as hiperligações que esta contiver, navegando para cada uma e baixando também o seu conteúdo, à semelhança do *crawler Python* anteriormente referido. A diferença está em não estarmos a guardar esta informação, servindo apenas de prova de conceito. Tudo é mostrado na mesma página ao utilizador.

Foram utilizadas sessões para não misturar resultados e permitir que cada utilizador tenha as suas próprias definições e esteja obrigatoriamente identificado (*"log in"* feito*)*. O fuso horário não é opcional e está definido como português. O objectivo é que a hora seja sempre a mesma independentemente do sítio ou da hora local. É utilizado um ficheiro de nome url_to_absolute que basicamente é um pequeno repositório de pequenos scripts que permitem partir o caminho do domínio para facilitar a navegação e impedir saltos para outros domínios se assim o pretendermos. Entre outras coisas. $base é o nosso domínio ou sítio web de origem. Tudo começa neste sítio e nas hiperligações aqui existentes. Código no apêndice "Código PHP".

4.5.4 Arquitectura, métricas, distribuição de carga e evolução iKNOW

4.5.4.1 Evolução da estrutura e arquitectura iKNOW

Versão 0.1 - O projecto assentava em scripts interligados que usando o terminal, obtinham directamente as informações do site pretendido e as enviavam para uma base de dados local, que as guardava, de alguma forma correlacionava e por fim, exportava um ficheiro .html local. Não havia complexidade pois cliente e servidor estavam na mesma máquina.

Versão 1 - No início, todo o trabalho assentava em *Raspberry Pi,* tanto para o servidor como para os clientes. Mas observou-se diversos problemas com os *Raspberry* servidores:

- eram lentos a disponibilizar conteúdos WEB a diversos clientes simultaneamente. Isto era devido às pesquisas e processamento que tinha de efectuar em bases de dados, motor PHP, e à fraca taxa de leitura que o cartão de memória fornecia (*e foram utilizados sempre os de classe mais rápida…*);
- devido ao alto processamento exigido aos *Raspberry Pi* como servidores, era relativamente fácil criar DOS/DDOS. Bastava pedir diversas páginas num curto espaço de tempo;
- corromperam-se uma série de cartões de memória. Sem explicação "científica" que explique o porquê, os servidores tinham uma alta taxa de avarias nos cartões. Sem certezas, sempre se atribuiu este facto à grande quantidade de *inputs/outputs* realizados ao cartão, à carga contínua que os cartões de memória estavam sujeitos e à pouca preparação destes para estes fins;
- a velocidade, nas versões 1, 2 e 3 era razoável, mas insuficiente devido à sua arquitectura de *hardware* partilhar o mesmo *bus* físico para USB e placa de rede. Com o mais recente Raspberry Pi 4, verificou-se uma grande melhoria na capacidade de resposta mas ainda assim, o problema do cartão SD manteve-se.

A arquitectura utilizada: a tradicional arquitectura cliente-servidor.

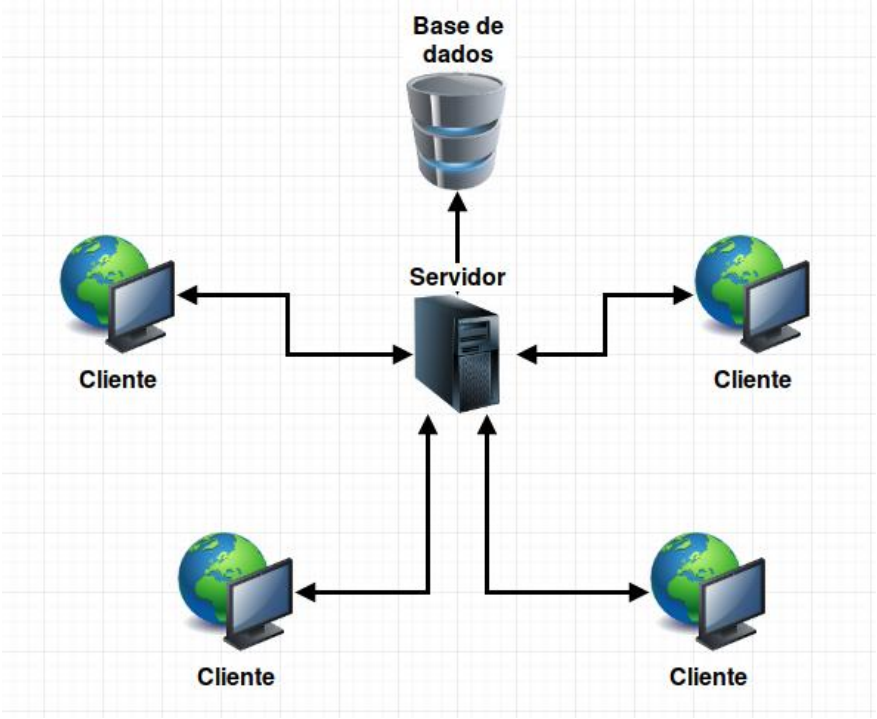


Figura 45 - Versão1. Arquitectura cliente-servidor mais comum

Versão 2 - Para tentar colmatar estes problemas, foi criado um balanceador de carga (também *Raspberry Pi*) que melhorou significativamente a velocidade das respostas, mas ainda assim não resolvia o problema.

Versão 3 - Após as considerações acima, optou-se por uma solução mais convencional e testada: a nova solução passou por trocar os *Raspberry* nos servidores Web e balanceador, mantendo-se nos clientes as bases de dados. Pelo que foi testado, os *Raspberry* não apresentaram quaisquer problemas como balanceadores. Nas bases de dados até ao momento ainda não se danificaram ou corromperam cartões de memória, mas devido ao elevado número de leituras e escritas, estará para breve…

Fica uma descrição e uma imagem (imagem 46) da solução:

- clientes iKNOW *raspberry* utilizam um ip fixo (potencial ponto de falha já que não usando DHCP, podem perder conectividade, a solução foi usar também a rede TOR) com que contactam o servidor, enviando e pedindo dados.
- servidor actua como *firewall* e balanceador de tráfego, dirigindo o tráfego para o servidor web com menos carga[168]. Servidor Linux com *Haproxy*, *firewall* e *log* de rede. (Poderia colocar-se aqui um IDS/IPS a registar a entrada de tráfego.);
- dois servidores web com *Nginx* e *PHP*, recebem e entregam os pedidos do balanceador, e fazem a comunicação com o *VIP* do *cluster* de base de dados *Mariadb/MySQL*;
- três bases de dados *MariaDB* em três equipamentos, utilizando o *cluster* Galera. O cluster exige mínimo de três máquinas de BD, fazem a replicação activo-activo entre si e garantem sincronização.

[168] Está a ser utilizado o modelo de balanceamento *round-robin*

A imagem 46, mostra o iKNOW, na sua terceira versão (2.0). Uilizada actualmente em "produção" (imagem simplificada[169]):

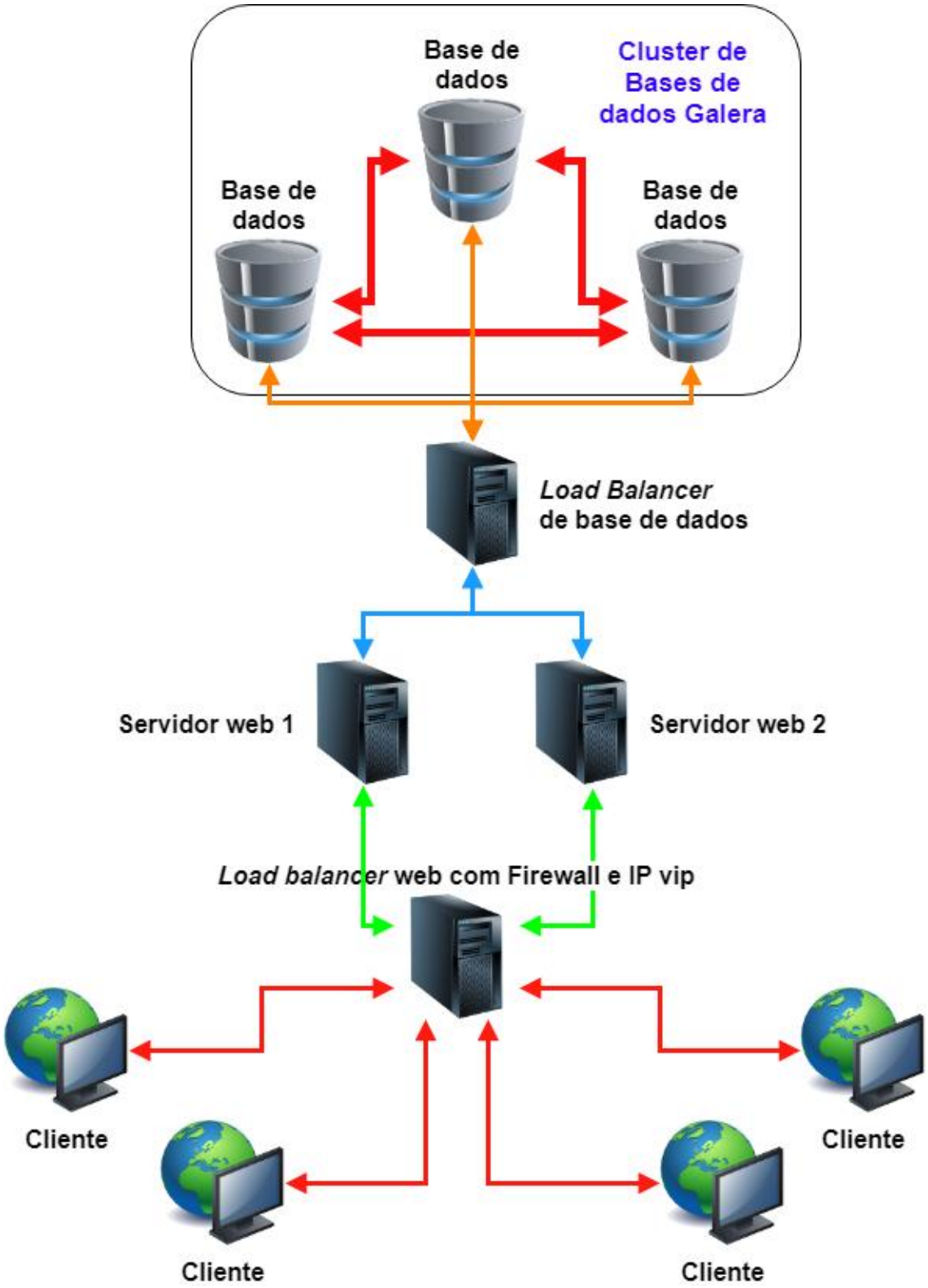


Figura 46 - iKNOW, versão 3, arquitectura melhorada e utilizada

[169] Imagem simplificada, pois está em execução a introdução de um segundo balanceador *web* inicial, para garantir a disponibilidade caso o balanceador "primário" falhe

A arquitectura aqui apresentada responde à necessidade e desafios a que nos propusemos, embora possam ser melhorados, especialmente se futuramente se conseguir integrar as características operacionais do iKNOW na sua primeira versão.

Dividindo a arquitectura pelos seus diversos módulos, temos diversos equipamentos, especializados em fazer pequenas coisas, o que permite a sua fácil substituição em caso de necessidade e redundância, caso se verifique a necessidade. Será agora explicada resumidamente cada um dos módulos.

Embora não sendo um sistema distribuído puro, incorpora algumas das suas funcionalidades, nomeadamente, o partir das tarefas grandes em pequenas, e o processamento destas, distribuído pelos clientes que as pedirem.

4.5.4.2 Servidor WEB, *scripting* PHP e balanceamento de tráfego

Foi escolhido o *NGINX* como servidor *WEB* por diversas razões, mas sobretudo pela sua maior "contenção" em consumo de memória e processador (foi também equacionado *Apache*. Estas razões foram essenciais visto que iria ser utilizado inicialmente apenas um servidor *Raspberry*, e tínhamos em mente, a utilização de muitos clientes (*raspberrys* como *crawlers* automáticos), ao mesmo tempo que utilizadores (humanos).

É um também servidor muito rápido nas respostas aos pedidos *web*, está bem integrado com a linguagem escolhida para *scripting* e intermediação *Web* (*html*) e a base de dados *MySQL*/*MariaDB*.

A carga que se pensou que iria estar sujeito era grande, por isso a escolha destes elementos assentou na leveza e na capacidade de resposta.

A linguagem de *scripting* PHP é conhecida do autor, tem capacidades únicas de interligação com a base de dados seja qual for a escolhida, tem bom suporte, existe em todas as distribuições Linux e tem uma excelente capacidade de resposta, especialmente a recente versão 7.x

Considerou-se que num caso real de negócio, o servidor não deveria ser um *Raspberry*, mas uma máquina ou máquina virtual dedicada. Não demorou a verificar que o *Raspberry* respondia com lentidão quando lhe eram exigidas páginas com processamento PHP intensivo, por exemplo, ler na BD, os registos referentes ao tempo de disponibilidade dos seus clientes/sondas. Mais, 4/5 clientes a pedir uma página em concreto (página PHP disponibilidade dos clientes) permitiu fazer um *DoS*, fazendo com que o servidor não tenha sido capaz de responder.

Como tentativa de resposta a estes problemas foi pensado e criado um *cluster* de servidores web, a responder como se fossem um, utilizando para este efeito, um balanceador de tráfego *HAProxy*.

4.5.4.2.1 Balanceador de tráfego

O balanceador de tráfego, como referido, foi o *HAProxy*[170]. O balanceador faz a partilha de carga entre os diversos servidores web. Optou-se pelo método *round robin*,

Foram utilizados dois balanceadores ou *network load balancers*, para distribuírem:

- LNB 1: pedidos/carga web
- LNB 2: pedidos e respostas a bases de dados

Colocado fisicamente na rede entre os clientes e o servidor web, distribuem a carga (acima descrita) para os servidores, fazendo assim com que cada servidor tenha menos carga, sendo mais rápidos na resposta e ao mesmo tempo, tenham mais tempo para as suas operações com a base de dados. Para um *raspberry*, este alívio de carga pode fazer a diferença entre funcionar e sofrer um DoS.

Quantos mais servidores web e bases de dados existirem, melhor será a distribuição. Mesmo com 2 servidores, foi relativamente fácil, sobrecarregá-los com duas páginas web muito intensas em termos de código e processamento PHP e Base de Dados. Novamente, foi utilizada a página de disponibilidade dos clientes (sondas), que é um ficheiro que lê TODOS os registos na BD por minutos que os *Raspberries Pi* estiveram online: centenas de milhares.

Nota: no processo inicial não foi utilizado balanceador nas bases de dados, porque o *cluster* estaria em modo activo-activo e portanto, cada servidor *web* escreveria para o seu próprio *MySQL/MariaDB*, e este depois faria a sincronização com todos os outros servidores de base de dados *MySQL/MariaDB*, através do cluster *Galera*. Vantagem: configuração inicial muito mais simples. Desvantagem: precisava sempre no mínimo de 3 servidores *web* cuja carga seria muito mais elevada porque fariam o *frontend*, *scripting*, operações de base de dados e ainda a sincronização entre estas.

4.5.4.2.2 Cluster de Base de dados: escolha, configuração e avaliação

A escolha da configuração, *cluster* e base de dados recaiu no *MariaDB*, uma base de dados relacional, em detrimento de outras bases de dados não relacionais (*NoSQL*) como o *MongoDB* ou o *Cassandra*. Ambas chegaram a ser equacionadas e testadas, pois são ambas muito utilizadas em clusters de *Big Data*, são resilientes e confiáveis. Mas o *cluster MariaDB/MySQL/Galera* permitia algumas características únicas e preciosas como se verá em vantagens e desvantagens.

Vantagens:

- **Instalação**: a instalação do *MariaDB* é simples de instalar. O *MongoDB* também, já para Cassandra teriam de ser compiladas algumas coisas;
- **Configuração**: o *Mariadb* tem bibliotecas muito testadas e prontas a usar com a linguagem de programação usada no site, o *PHP*. O *MongoDB* e o *Cassandra* exigem mais configuração;
- **Continuidade de serviço**: queremos que em caso de falha, os serviços continuem a funcionar, independentemente do ponto de falha. O *cluster Galera* permite isto: independentemente do membro que falhe, tudo continua a funcionar. Caso volte a ficar activo, as comunicações são automaticamente restabelecidas e os dados sincronizados;

[170] http://www.haproxy.org/

Há aqui um "empate" entre *Cassandra* e *Galera*.
- **Sincronização:** queremos uma topologia activo-activo, com sincronização. Quere-se também que qualquer membro do cluster possa receber e transmitir os dados. A utilização da Galera como base de cluster, permite sincronização numa topologia activo-activo. Pelo que foi referido, também aqui ganha o *Galera*, embora haja um empate com o *Cassandra*;
- **Facilidade de utilização:** o Galera e *Mysql*/MariaDB já eram conhecidos do autor, já têm bibliotecas para funcionar com a linguagem de programação Python e PHP, é simples e rápido de instalar, além de todos os sistemas Linux recentes terem esta ferramenta;
- **Transaccional**: se a operação não tiver sido concluída com sucesso, ela não é executada. "Ou é bem feito, ou não é feito". Não há operações a meio ou perdidas.

Desvantagens: Foram testadas outras bases de dados como as acima referidas (*Mongo* e Cassandra) para *Big Data*, mas a capacidade de armazenamento, processamento e RAM do *Raspberry PI* (onde se quer usar a base de dados), não permitem realmente a possibilidade de fazer uso de *Big Data* no *Raspberry*. Portanto, a partir do momento em que o *Galera* começou a ser possível no *Raspberry* (a partir da inclusão do *MariaDB* 10.1, algo muito recente), a sua escolha foi óbvia. Os resultados obtidos são muito bons, pese embora a quantidade de cartões SD que sofreram avarias terem sido muitos. Deduz-se que o problema tenha sido desgaste pelo excesso de leituras e escritas.
Também na rapidez da manutenção, o Galera é simples, mas quando temos o sistema a funcionar, era mais simples o Mongo, pois podemos acrescentar dados e novas colunas sem estar a declarar tudo à base de dados.

Clientes: utilizados neste projecto (*Raspberry Pi*), foram sempre pensados como sondas, com o objectivo de dividir o trabalho, quer em processamento quer na geografia. Devem estar geograficamente divididos, ser leves, executar tarefas simples, como é a descarga de páginas web, e ler e escrever no/do servidor ordens/resultados. Os clientes não têm informação completa de quem pede ou do porquê. Apenas têm uma lista de ordens e escrevem depois o resultado na base de dados. A comunicação cliente-servidor ainda não é cifrada no envio de dados para a base de dados. Mas é cifrada na comunicação https e está criado mas ainda não a ser utilizado, o uso de comunicações cifradas servidor-clientes via biblioteca *Python Paramiko*.

Na figura 47, vê-se quatro terminais a correr simultaneamente em quatro sistemas diferentes. Pode verificar-se que: em um dos terminais é criada uma base de dados; em outro é criada uma tabela; outro é inserido informação na bd/tabela e outro é mostrado o que aconteceu. Todas se afectam mutuamente replicando e sincronizando todos os comandos.

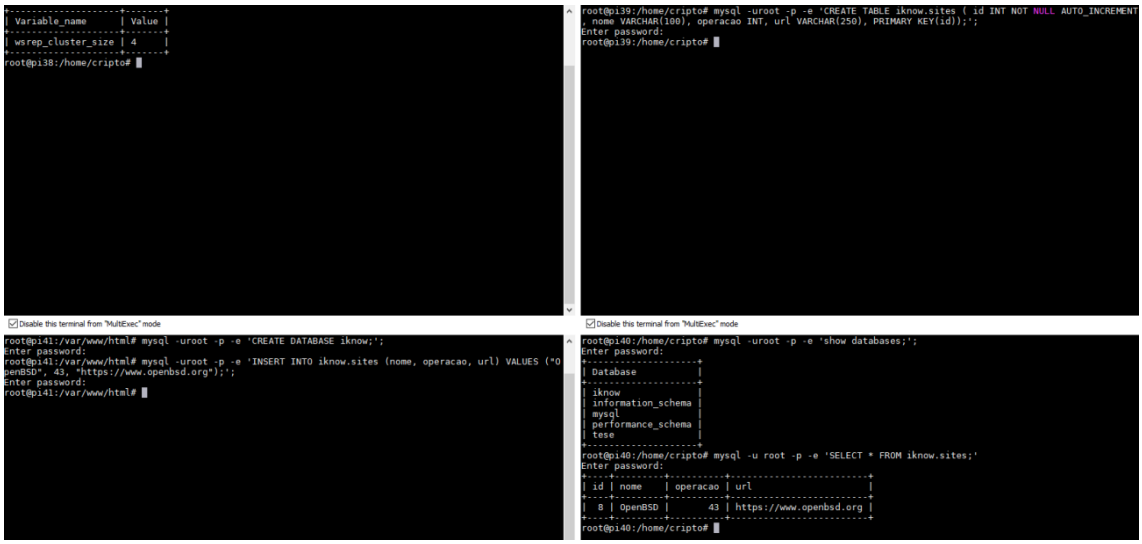


Figura 47 - Cluster de base de dados Galera a sincronizar e replicar informação

4.5.5 Instalação e configuração de clientes e servidor

A instalação dos clientes e servidor, foi pensada para ser automatizada, de forma a simplificar e agilizar este passo tão importante, trabalhoso, moroso e ("possivelmente") inibidor para muitos.

Cliente: resumidamente, é pedido ao utilizador que queira colocar o seu *Raspberry* (ou sistema Linux, com algumas alterações ao código), que corra um comando como *root*, que por sua vez, fará a descarga do ficheiro e fará toda a instalação, quase automaticamente (sendo pedido de vez em quando ao utilizador, uma ou outra informação). Depois de instalado, é pedido ao utilizador que edite o seu *crontab* como utilizador *root*, de forma a automatizar depois o iKNOW. Por uma questão de facilidade, também foi clonada a imagem original do cartão SD para ser replicada, sendo a desvantagem o tamanho (8gb).

Configurações: por uma questão de segurança, todas as configurações nome de utilizador/passe estão num ficheiro de nome "dados.py" (é importado por todos os módulos e precisamos ter cuidado com este ficheiro e se quisermos fazer *upload* na aplicação, devemos mudar as credenciais e a localização do ficheiro para local inacessível do servidor *web*).

Servidor: à semelhança do cliente, a instalação do servidor *web*, linguagem de *scripting PHP*, etc, são automatizadas, com perguntas esporádicas. Depois de instalado o *script* é necessário inserir os dados iniciais de configuração na base de dados e copiar para a pasta do servidor *web*, os documentos iniciais do iknow (páginas *html*, *scripts*, imagens, etc). Necessário ter portos 80, 443 e 56789 (*ssh* em porta alterada) abertos. O porto 80 é só para testes porque vai utilizar certificados SSL (*https* - 443) nas comunicações da plataforma.

Base de dados: para funcionar a plataforma servidora devemos instalar uma base de dados de nome MySQL ou MariaDB. Após instalação da mesma, devemos importar um ficheiro que contém todas as tabelas e informações que permitem ao IKNOW funcionar. Os clientes não

precisam de base de dados servidor mas precisam do cliente para enviar a informação para o servidor. Resumidamente é necessário criar uma base de dados e importar a nossa.

Serviços TOR - cliente: para aceder às páginas e rede TOR, é instalado o software e serviço aquando da instalação do cliente. Correm automaticamente ao iniciar e permitem o acesso e comunicações na rede TOR. Este serviço no cliente permitirá futuramente comunicar de forma segura entre servidor e cliente, impossibilitando a localização do cliente por terceiros que tenham acesso malicioso ao iKNOW. Futuramente, os *updates* também vão decorrer via TOR.

Serviços TOR – servidor: o servidor web funciona caso se deseje, com acesso via rede TOR. Para isso precisa ter este serviço activo, e ainda, estar configurado para fornecer serviços web. Esta instalação é automática no ficheiro de instalação do servidor, incluindo, a criação de uma página web que indica qual o endereço web na TOR.

Balanceadores e *cluster Galera*: no âmbito deste projecto não foram construídos *scripts* para os balanceadores web e/ou base de dados. No entanto a instalação e configuração dos HAProxy para os dois serviços, são bastante simples e estão muito difundidos na Internet. Também a instalação do *cluster* Galera, embora não tão simples e directo, é passível de ser feito seguindo um dos muitos e bons manuais na Internet. São necessários para o *cluster Galera*, um mínimo de três equipamentos.

Configuração do *crontab* como super-utilizador (comum a ambos os instaladores) de forma a correr automaticamente o script de cinco em cinco minutos. O código a ser inserido no *crontab* (*não esquecer as permissões necessárias e colocar o script com privilégios de execução*):

***/5 * * * * /home/pi/python-mysql/pi_actualizar_servidor.py**

O instalador automático abaixo, pode estar desactualizado e deve, portanto, ser sempre utilizada a última versão. Serve, contudo, de referência e prova de conceito. Ficaria com o seguinte código (que se coloca abaixo por estar comentado e ser reduzido):

```
INSTALADOR AUTOMÁTICO
#====================== iknow-instalar.sh
clear ;wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/iknow-instalar.sh -O ./iknow-instalar.sh
sudo chmod +x iknow-instalar.sh
clear
./iknow-instalar.sh
echo "Vamos fazer o download do instalador e correr o mesmo.."
wget --user-agent="iKNOW-instalador/1.0 (X11; U; xpto i686; pais; rv:1.0.0.0) Gecko/2008092416 Firefly/1.0.0"
"http://www.gualdimpais.dtdns.net/ficheiros_cliente/instalador.sh" -O "instalador.sh"
echo "Download concluído... Vamos dar permissoes.."
sudo chmod +x ./instalador.sh
echo "Instalador com permissoes, execute: sudo ./instalador.sh"
sudo ./instalador.sh
#wget -r -H -nc -np -nH --cut-dirs=1 -e robots=off -l1 -i ./itemlist.txt -B 'http://archive.org/download/'
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/instalador.sh
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/lista.txt
wget -i lista.txt -B 'http://www.gualdimpais.dtdns.net/ficheiros_cliente/
echo "====================== iknow-instalar.sh – fim"
```

Já o instalador manual, permite ter um pouco mais de controlo e aprender um pouco como as

coisas estão a ser feitas. Todas as operações estão comentadas:

1-Transferir o sistema operativo RaspBian Light (modo consola, sem *interface* gráfica)
2-Transferir o sdFormatter
3-Formatar o cartão SD com o sdFormatter
4-Transferir o win32diskImager
5-Usar o win32diskImager:
5.1-Seleccionar a imagem do RaspBian (ficheiro.img, se estiver .zip tem de descompactar primeiro.)
6-Criar um ficheiro de nome "ssh" (sem terminação)
6.1-Copiar os 2 ficheiros para o disco BOOT (o cartão sd ficou com esse nome após transferida a imagem)
Os 2 ficheiros criam acesso SSH após o boot, e indicam qual o ip que o raspberry irá usar.
Podem ser editados antes ou depois de ser transferidos para o cartão SD

6.2 - ALTERAR O IP PARA FIXO
Alterar no cartão o ficheiro cmdline.txt acrescentando no final da linha existente (algo como "*dwc_otg.lpm_enable=0 console=serial0,115200 console=tty1 root=/dev/mmcblk0p2 rootfstype=ext4 elevator=deadline fsck.repair=yes rootwait"*), o seguinte texto, adaptado à realidade: "ip=192.168.1.204::192.168.1.1:255.255.255.0"

7-Ligar o raspberry. Pode já aceder ao raspberry utilizando:
Login:pi
Passe:raspberry
8-Mudar a palavra-passe:
passwd

8-Expandir sistema para ocupar todo o tamanho do cartão. Mudar o nome do sistema e
sudo raspi-config
- Expandir cartão
- Mudar hostname. Exemplo: pi204 (nome+ip)
- Mudar boot para consola (embora já o esteja a fazer com a versão light)
- Mudar 64 megas de gráfica para 16 ou 32 (não vamos usar interface gráfico)

PARA OS PASSOS SEGUINTES É PRECISO INTERNET

9-Depois de autenticar, vamos actualizar tudo (demora um pouco mas é essencial):
sudo update -y
sudo upgrade –y

10-Instalar e configurar um firewall (últimas versões do Raspbian já vêm com o UFW pré-instalado)
sudo apt-get install ufw -y
sudo ufw default deny incoming
sudo ufw default allow outgoing
sudo ufw allow ssh
sudo ufw allow http
sudo ufw allow https
sudo ufw allow 9050
sudo ufw allow 9150
sudo ufw allow 3306
sudo ufw enable
sudo ufw status numbered

11- Instalação do serviço TOR. Útil para aceder à rede TOR e fornecer serviços
sudo apt-get install tor –y

13-Editar:
sudo /etc/tor/torrc
Alterar:

```
HiddenServiceDir /var/lib/tor/hidden_service/
HiddenServicePort 22 127.0.0.1:22

14-Iniciar serviço e Reiniciar para o tor criar o ficheiro hostname
sudo service tor start
sudo reboot

15-Obter o endereço TOR (que depois deve ser partilhado para aceder ao servidor via TOR), criar pasta para o
projecto e dar as permissões necessárias para correr

Echo "Endereco TOR do servidor"
sudo cat /var/lib/tor/hidden_service/hostname
==========instalador.sh
clear
echo "-----------------------------------------"
echo "                Instalador do iKnow                "
echo "-----------------------------------------"

echo "Se o utilizador nao for o pi, tem de alterar estes caminhos..."
echo "Precisa de ter privilégios sudo.."
echo ""
echo ""
echo ""
echo "Criando pasta /home/pi/iknow"
sudo mkdir /home/pi/iknow
sudo chmod 777 /home/pi/iknow
echo "Entrando na nova pasta.."
cd /home/pi/iknow

echo "A fazer download dos ficheiros necessarios:"
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/instalador.sh -O /home/pi/iknow/instalador.sh
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/actualizar.sh  -O /home/pi/iknow/actualizar.sh
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/dados.py  -O /home/pi/iknow/dados.py
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/instrucoes_crontab.txt  -O
/home/pi/iknow/instrucoes_crontab.txt
wget http://www.onossoiknow.dtdns.net/ficheiros_cliente/pi_actualizar_servidor.py  -O
/home/pi/iknow/pi_actualizar_servidor.py
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/temp_bash.sh  -O /home/pi/iknow/temp_bash.sh
wget http://www.gualdimpais.dtdns.net/ficheiros_cliente/versao.txt  -O /home/pi/iknow/versao.txt

echo "Vamos dar privilegios de execucao aos ficheiros necessarios:"
sudo chmod +x instalador.sh
sudo chmod +x actualizar.sh
sudo chmod +x pi_actualizar_servidor.py
sudo chmod +x temp_bash.sh

echo "Instalar conector de base de dados mysql-mariadb-python e todas as bibliotecas necessarias"
sudo apt-get install python-mysqldb -y
==========instalador.sh - fim
```

Testar: Se o comando "./home/pi/python-mysql/pi_actualizar_servidor.py" não devolver erros, está a funcionar. Devido ao *crontab*, estas acções serão executadas de cinco em cinco minutos.

4.5.6 Análise dos alvos – Fase do planeamento

Esta fase segue o ciclo OSINT. Antes da recolha deve ser definido claramente o que se pretende e só depois avançar para a recolha em si. Dependendo do site que se pretenda criar operações,

a ferramenta iKnow pode ou não funcionar e devolver os resultados pretendidos. Se não der, tem de ser alterado o código para responder à forma como o site foi criado.

Os sítios (páginas web, ficheiros, ou outros) de onde se pretende obter informação devem ser previamente (se possível), estudados. Deve ser feita uma análise de como a informação está estruturada (se estiver), e o que importa retirar. Por exemplo, um sitio web de notícias poderá ter algum mecanismo de *feeds.* Se tiver pode poupar tempo e recursos tanto a nós, como ao próprio sitio web. O sítio web da Europol[171] tem uma zona onde é possível ver a lista dos criminosos em fuga[172], mas tal informação não é obtida facilmente através de um simples download da página, portanto torna-se necessário ler atentamente como está estruturada, onde está o que interessa e de que forma as imagens estão a ser "escondidas". Depois de compreendido o funcionamento e de escrita uma rotina para o caso do sitio web, as informações vão fluir até que alguém mude novamente o funcionamento do sitio web ou o nome das coisas (*infelizmente, esta situação acontece cada vez com mais frequência em todo o tipo de sítios web...*). O estudo analítico de cada site permite a obtenção pura da informação e não o todo (que como já vimos, é prejudicial).

Exemplo: a imagem abaixo é o que se obtém quando se tenta baixar a imagem através do sítio *web* da *europol – most wanted*. Não há imagem. Foi criada uma protecção para evitar que *crawlers*, obtivessem automaticamente estas imagens (mas como veremos, com o iKNOW, é possível).

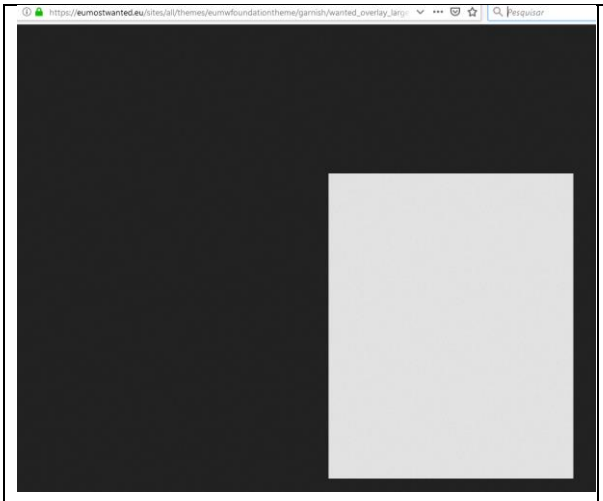


Figura 48 - Imagem dos mais procurados da Europol - com protecção. Nada se vê

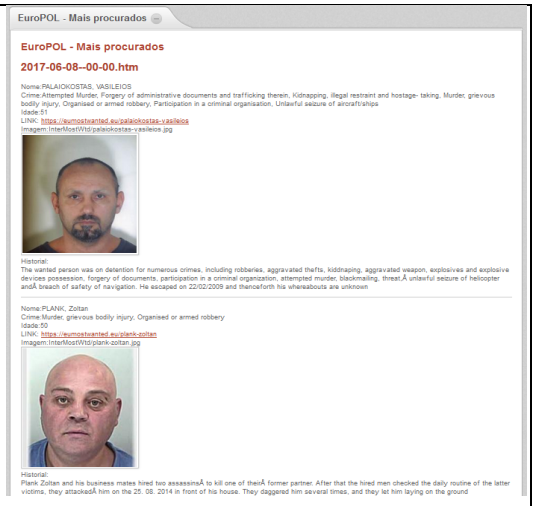


Figura 49 - Imagem recolhida e visivel, através do iKNOW

4.5.7 Técnicas e limites de *crawling/scraping*

As buscas automáticas e de alto volume de dados ou de *harvesting* geram bastante tráfego que pode ser facilmente detectado e mesmo que não o fosse, é eticamente reprovável pois estamos

[171] https://www.europol.europa.eu/
[172] https://eumostwanted.eu/

a sobrecarregar o site com pedidos, podendo mesmo causar uma negação de serviço capaz de o colocar offline ou inoperacional durante muito tempo, com os prejuízos daí decorrentes.

Foram encontradas e colocadas em prática as seguintes soluções:

- Simulação de *browser* recorrendo a *user-agents* que dão o nome e versão do *browser*, e *referrer* (site de onde veio o visitante). Todas estas variáveis são aleatórias, mas baseados em elementos reais (ex: *browser* Firefox 52 no Windows 7 é dado através da informação "Mozilla/5.0 (Windows NT 6.1; WOW64; rv:52.0) Gecko/20100101 Firefox/52.0");
- Abertura de *sites*, simulando um clique humano, demorando um tempo aleatório;
- Utilização de métricas: durante um dia, só são feitos x pedidos a um determinado site, independentemente da quantidade de sistemas em funcionamento (exemplo: apesar de termos 25 máquinas, no máximo o servidor só teria 100 pedidos. E não 100 pedidos vezes 25 máquinas, o que daria 2500 pedidos por dia);
- O limite de pedidos é ultrapassado sem grande prejuízo se tivermos vários projectos, em que os equipamentos vão variando o que vão fazendo sem sobrecarregar nenhum servidor e tendo em conta que os pedidos são distribuídos por todas as máquinas e não por apenas uma (alguns servidores já bloqueiam ou pedem um desafio CAPTCHA para verificar se é um humano ou uma máquina a pedir informações);
- Se um determinado site bloqueia um IP tendo em conta a sua origem (nacionalidade por exemplo), é possível contornar utilizando uma VPN, a rede TOR ou colocando uma máquina nossa naquele país… (sonhando que tal é uma possibilidade se o site e a ferramenta começassem a ser utilizados por muita gente… mas possível). Exemplo: o site da *Al Jazeera*[173] (notícias árabes) em terreno americano tem ip daquele país;
- Obtenção de dados e coordenadas GPS a partir das imagens obtidas através do *crawl.* Estas coordenadas são armazenadas em local diferente na base de dados, mas associado à operação que lhe deu origem;
- A colocação de sistemas *iknow* em diversos sítios do país e do mundo possibilitaria também o seguinte: compra de bilhetes online mais baratos, fugindo às manhas dos sites de hotéis que detectam a localização do ip do visitante do site, conforme imagem 50 (*que não foi possível saber a fonte*).

Figura 50 - preços variam consoante a localização geográfica

[173]https://www.aljazeera.com/

Qualquer das técnicas acima tem em conta, uma utilização ética dos recursos dos sites que se tenta obter informação. Um utilizador/*hacker* sem escrúpulos pode simplesmente bombardear o site de tal maneira com pedidos, que todos os outros utilizadores irão sentir o site mais lento e em último caso, vai conseguir tirar o site ou serviço do ar, criando uma situação dos conhecidos ataques de *DoS/ DDoS*.

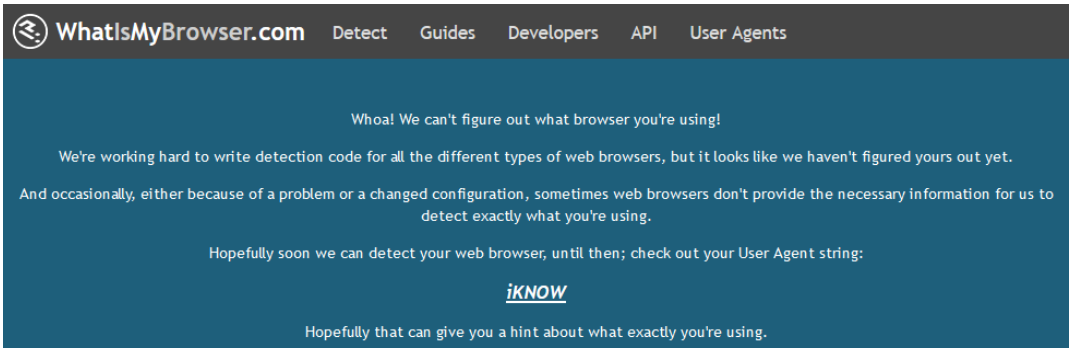
4.5.8 Exemplificação prática de algumas funcionalidades

Para se obter informações de páginas web, foram criadas e testadas várias soluções, sendo que praticamente todas elas simulam utilizadores humanos. Utilizando PHP na fase inicial simulamos o navegador.

Visitando o sítio web[174] Com o navegador original:

Com o navegador forjado, em PHP, unicamente a dizer que é *Firefox*, funciona e portanto é "forjar" também:

A título de exemplo, simular um navegador fictício, não traz resultados convincentes, e é facilmente identificado por "qualquer" pessoa:



O iKNOW utiliza diversas técnicas para simular não só o *user-agent* mas também o comportamento humano (demora nos cliques por exemplo), dificultando a sua detecção quer por *firewalls*, ou por analistas de segurança.

[174] https://www.whatismybrowser.com/

Conteúdos "obscurecidos": Utilizando o iKNOW é possível por exemplo ultrapassar algumas barreiras básicas (não de segurança mas de obscuridade) que os criadores dos sites desenvolveram para que seja possível ver mas não copiar. Por exemplo, o site da Europol na zona de procurados. Com o iKNOW podemos baixar a lista de pessoas procuradas, as "malandrices" de que são acusadas e a foto (se tentarmos ver ou baixar a foto directamente do site, apenas baixamos uma imagem diferente). Note-se que embora estejamos a rodear um mecanismo de obscuridade da foto, ela está lá à disposição de qualquer um, publicamente portanto é OSINT. Em baixo, a feed criada pelo iKNOW. O *crawler* e *parser Python*, funcionam acedendo ao site, baixando o conteúdo e aplicando uma série de filtros que foram criados de forma manual para que se obtenha apenas a informação desejada. Estes filtros para funcionarem correctamente têm de ser feitos site a site, se o site não for genérico ou contiver tabelas e outras informações de forma menos "*standard*". O caso abaixo foi trabalhoso inicialmente no estudo das *tags,* mas depois disso, ficou a funcionar muito bem (caso os autores do site da Europol alterem o nome das *tags* ou o comportamento, o crawler/parser deixam de funcionar...).

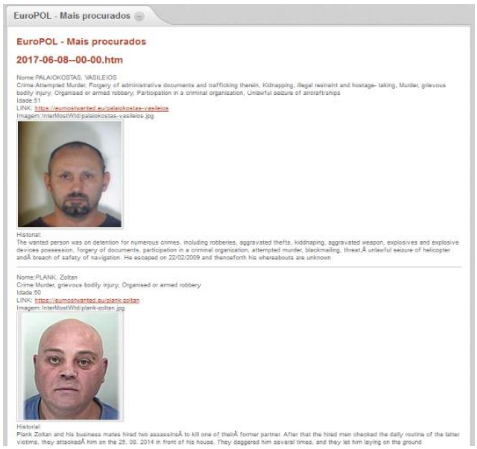


Figura 51 - FEED iKNOW de procurados Europol

Conteúdos escondidos: Utilizando a ferramenta iKNOW, é também possível obter notícias de alguns sites que obrigam o utilizar a clicar "ler mais". Com alguns sites até é possível ler a noticia que apenas está disponível (*em teoria*) como *Premium*, mediante pagamento. Abaixo estão duas imagens: na primeira, o aviso que não podemos aceder ao conteúdo porque já visualizámos o limite gratuito; na segunda, todo o conteúdo é obtido pelo iKNOW. Nenhuma protecção é ultrapassada, bastava ver o código-fonte: a informação está lá mas não acessível directamente.

Figura 52 - obtenção de informação oculta

Meta-dados de imagens: enviar uma imagem e verificar se tem coordenadas. Tendo, obter a localização num mapa.

Localizar produtos furtados: é possível utilizar *plugins* iKNOW para pesquisar em sites e fóruns onde se vendem produtos. Como por exemplo OLX. O funcionamento deste *plugin* especifico é muito simples. Permite que nos autentiquemos com o nosso script, e de seguida, procure como se se tratasse de uma pesquisa manual. Todos os resultados podem ser exportados no formato *html* e visualizados num navegador web normal. (*Ainda não se programou a colocação dos resultados numa base de dados nem configurou para que apenas recebamos resultados úteis, sem publicidade ou scripts secundários daquele site*).

Pesquisar por termos associados a criminalidade e terrorismo: utilizando as potencialidades da rede TOR, podemos fazer pesquisas a sites comuns na Internet, protegidos pela anonimidade fornecida tanto por VPN como por endereços obtidos na rede TOR.

O iKNOW permite fazer pesquisas tendo como partida, um sitio web. Utilizando simples *queries* como a referida acima para produtos furtados, é possível utilizar também termos *regex* e *fuzzing* para pesquisar nomes, locais, equipamentos e outros, relacionados com criminalidade e terrorismo.

O funcionamento do que já tinha sido criado era bastante simples. Um *script* é executado tendo como partida um ou mais sítios web. Esse script vai varrer todas as hiperligações que encontre e se encontrar mais termos, na lista de termos que pretendemos encontrar, regista, obtém o site e insere na base de dados a quantidade de termos, o local e a hora. O *fuzzing* serve para conseguir obter termos que possam estar ligeiramente mascarados como por exemplo, “te..rror..ismo” ou “b i n la den” e semelhantes. Actualmente existe o *script* mas falta integrar na versão 2.0. O resultado seria depois incorporado não num relatório OSINT, mas num ficheiro pdf com tudo o que tinha o site.

A classificação de um site ter ou não interesse é baseado em métricas muito simples: se aparecer um ou mais termos, e se esses termos se repetirem, são somados. Mais de “x” termos tem interesse e merece o download da página, a sua inserção, e a sua posterior conversão em PDF e envio.

Líbia

O Estado Islâmico já tem um pé na Líbia

Cidade do Leste tornou-se bastião dos jihadistas do EI num país sem autoridade de Estado e numa situação caótica.

IMED LAMLOUM AFP - 1 de Dezembro de 2014, 10:01

Uma coluna com bandeiras do grupo jihadista EI perto da cidade líbia de Derna REUTERS

A ideologia extremista do grupo Estado Islâmico (EI) atraiu simpatizantes nos meios radicais na Líbia, onde a cidade de Derna, no Leste, transformada em "emirado islâmico", se tornou um feudo dos partidários do EI, segundo os especialistas.

Alguns observadores ocidentais consideram já esta cidade de 150 mil habitantes, histórica praça-forte dos radicais islâmicos no Leste líbio, como o terceiro ramo do EI na África do Norte, depois dos grupos Jund al-Khilafah, na Argélia, e Ansar Bayt al-Maqdis, no Egipto.

O dirigente do EI, Abu Bakr al-Baghdadi, vangloriou-se recentemente num registo radiofónico da expansão do "califado", anunciando ter aceitado os juramentos de fidelidade dos jihadistas da Líbia, do Egipto, do Iémen, da Arábia Saudita e da Argélia. "O EI está em Derna. Isso está bem documentado. Não há qualquer dúvida", sublinhou o político líbio Othman Ben Sassi.

O grupo "beneficia da ausência de autoridade do Estado e de fronteiras porosas" num país caótico, onde dois parlamentos e dois governos disputam o poder num clima de violência e derramamento de sangue, lamenta este antigo membro do Conselho Nacional de Transição (CNT), que foi o braço político da rebelião contra Khadafi.

Há semanas que comunicados e fotos circulam em fóruns radicais sobre a adesão dos "jihadistas da Líbia" ao EI. No entanto, esses documentos não são subscritos por qualquer grupo e têm a vaga assinatura de "Estado Islâmico, Wilayat Barqa" (província de Cirenaica, região leste).

Os Estados Unidos expressaram recentemente a sua "inquietação" com base em "informações segundo as quais facções extremistas violentas [na Líbia] prometeram fidelidade ao EI e procuram associar-se a ele", segundo o porta-voz do Departamento de Estado, Jeffrey Rathke.

Desde a queda do regime de Muammar Khadafi, em 2011, Derna e uma grande parte de Bengasi, segunda cidade do país, estão nas mãos de grupos radicais, em particular do Ansar Al-sharia, classificado como organização terrorista pela ONU.

"Conflito ideológico"

Em Abril, um ramo do Ansar Al-sharia anunciou ter instaurado a justiça de acordo com a lei islâmica em Derna. Denominando-se "Majlis Choura [Conselho Consultivo] dos jovens do islão em Derna", instaurou tribunais islâmicos e uma polícia religiosa.

Dezenas dos seus membros encapuçados e vestidos com uniformes militares desfilam regularmente em viaturas *pick-up*, armados de lança-*rockets* RPG e metralhadoras, empunhando a bandeira preta e branca dos jihadistas. O grupo divulgou em Agosto um vídeo da execução pública de um egípcio acusado de assassínio, no estádio da cidade, a primeira vez que isso aconteceu na Líbia.

Este grupo extremista não anunciou até agora formalmente a sua fidelidade ao chefe do EI e analistas falam em divisões nas suas fileiras. "Vários extremistas de Derna sentem-se atraídos pelo EI. Mas a maior parte dos chefes jihadistas na Líbia são antigos membros da Al-Qaeda e há um conflito ideológico entre os partidários do EI e os da Al-Qaeda", diz um especialista líbio, sob anonimato.

A ONU classificou este mês o Ansar Al-sharia em Bengasi e Derna como organizações terroristas devido aos seus laços com a Al-Qaeda do Magrebe Islâmico (AQMI). "Esta

Figura 53- Obtenção por termos-chave: Estado Islâmico

Relatórios OSINT semanais: enviar uma mensagem por Telegrama, colocar uma URL no site, automatizar a recolha de capas de jornais diários, enviar relatórios de forma cifrada para o email, criar relatórios com o slogan da nossa instituição, …. Tudo isto a partir de uma única plataforma, meia dúzia de cliques e configurável. (Ver Relatórios OSINT para ver exemplos.)

Motores de busca: foi criado um script para que o iKNOW navegue no Google, simulando um navegador web e obtenha informação, sendo possível saltar/"*crawlar*" o Google. No entanto, recentemente o Google criou uma série de protecções para que o utilizador não consiga através da ligação mostrada pelo Google, saber qual a página/domínio de destino *(é mau.. talvez muitos o façam, abusando sem limites ou consideração pelos outros, e criando dificuldades depois a todos...).*

Sítios web de venda de produtos: é possível adaptar o *crawler/scrapper* básico para que este verifique em diversas lojas de informática, o preço de produto A e o compare entre elas. Foi feito a nível de prova de conceito, um script que corre em linha de comandos e obtém o disco X do site da loja A e o mesmo disco X do site da loja B. Utilidade: verificar pelo produto mais barato.

Em baixo, um exemplo de como é feita a obtenção dos links via PHP e geração de página: é feito também através da simulação de um navegador/*browser,* neste caso baixando e mostrando as hiperligações existentes:

```
<html Content-Type: text/html; charset="utf-8">
<?php
#$url = "http://www.google.com";
#$url = "http://www.packtpub.com/learning-ext-js/book";
$ur l ='https://www.globaldata.pt/shop/discos-rigidos/sata-3-5.html';

$ch = curl_init();
$timeout = 5;
curl_setopt($ch, CURLOPT_URL, $url);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
curl_setopt($ch, CURLOPT_CONNECTTIMEOUT, $timeout);
curl_setopt($ch, CURLOPT_USERAGENT, "Mozillax/4.0 (compatible; MSIE 8.0; Windows NT 6.0)");
curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
```

```
curl_setopt($ch, CURLOPT_SSL_VERIFYHOST,false);
curl_setopt($ch, CURLOPT_SSL_VERIFYPEER,false);
curl_setopt($ch, CURLOPT_MAXREDIRS, 10);
curl_setopt($ch, CURLOPT_FOLLOWLOCATION, 1);
curl_setopt($ch, CURLOPT_CONNECTTIMEOUT, $timeout);

$html = curl_exec($ch);
curl_close($ch);

$dom = new DOMDocument();

# @ utilizado para esconder avisos derivados de tags html inválidos existentes
@$dom->loadHTML($html);

# percorrer todas as tags <a>
/*
foreach($dom->getElementsByTagName('a') as $link) {
	# Show the <a href>
	echo $link->getAttribute('href');
	echo "<br />";
}
*/
foreach($dom->getElementsByTagName('img') as $link) {
	# Show the <a href>
	//echo "<br/>".$url.$link->getAttribute('src');
	echo "<br />";
                        $imagem=$link->getAttribute('src');
                        echo $imagem.'<br/><img src="'.$imagem.'">';
}
?>
```

CAPÍTULO V – Avaliação

5.1 Objectivos

O iKNOW aquando do seu início (2014[175]) pretendia fazer algo que não era conhecido nem comum: um site central com um motor de busca próprio, com base em termos definidos por este, de forma mais precisa, com um limite de profundidade e com facilidade de utilização. Não é para ser um "Google" da Internet, mas apenas uma ferramenta que poupe trabalho e trabalhe autonomamente informando quando houverem notícias ou palavras-chave do que pretendemos. A ferramenta levou muitas alterações ao longo dos anos e de alguma forma, tornou-se cada vez mais complicada a sua manutenção, devido aos sites cada vez mais dificultarem o acesso à informação, incluindo o próprio Google.

Objectivos:

- Capturar dados OSINT, recorrendo a um conjunto de equipamentos, distribuídos geograficamente;
- Processar e analisar os dados obtidos;
- Difundir os resultados obtidos, quer seja em notícias concretas, que em gráficos, quer em relatórios;
- Copiar e reter dados *offline* numa base de dados pesquisável e disponível com toda a informação obtida. A informação guardada *offline* servirá para consulta e extracção e co-relacionamento de dados;
- Obtenção de dados do site, seja qual for o país e o tipo de codificação de caracteres que tenha. Por defeito esta informação é depois guardada em base de dados em UTF-8 e/ou em ficheiros *html;*
- Localizar cartões de crédito, termos e palavras-chave;
- Pesquisar em motores de busca nomeadamente no Google;
- Pesquisar na internet dita "normal" ou na rede *TOR;*
- Utilizar um *Web crawler* criado para o efeito, para pesquisar pelas fontes pretendidas, saltando de sítio web em sítio web conforme vá obtendo informações de relevo;
- Web *parser* para retirar informação;
- Obter localizações e dados GPS, através das imagens fornecidas, ou das imagens obtidas;
- Obter endereços ips dos visitantes, hora, data, identificação de navegador, etc, para efeitos de melhor fornecer o serviço de páginas web;
- ~~*Facebook*~~ - Anteriormente era possível pesquisar via API na rede social *Facebook*. Descontinuado devido a API ser de uso específico para utilizador e rede de sensores não poder usufruir de tal. Entretanto a API sofre modificações e tornou-se impraticável. Pouca utilidade comparativamente ao grau de dados recolhidos e dificuldade de execução, com exigências do *Facebook* de API por utilizador;

[175] Já lá vão cinco anos...

- ~~*Tweeter*~~ – pesquisa nos *tweets* de pessoas-alvo por hora de colocação de notícia e pessoas que seguiam. Objectivo era definir horas de actividade da pessoa e horas de descanso. Foi descontinuado;

Onde: nos *sites* que quisermos, com algumas restrições nomeadamente se forem fóruns ou *sites* que exijam autenticação (e do tipo da mesma). É possível organizar o iknow para se autenticar em *sites* específicos e depois fazer a busca (mas é mais moroso e tem de ser feito manualmente pois é altamente específico… mas possível).

Como: varredura temporal pré-definida por nós, no sítio web pretendido, seja na internet "normal" ou na rede *TOR*, com o guardar automático de toda a informação que tivermos definido. Pode ser definido que apenas se queira obter relatório se termo x ou y aparecerem no *site*.

Quando: com o espaço de tempo que definirmos ou apenas uma vez (as tarefas executadas pelos diferentes clientes *crawlers*, são feitas pela ordem de criação de operações, e verificadas minuto a minuto, e só param quando fizerem x saltos em sites ou não houver mais hiperligações ou informações a guardar).

5.2 iKNOW - Propostas de melhorias nos clientes e servidor

5.2.1 *Multithreading* nos clientes

As tarefas dos clientes dividem-se em pequenos e simples passos. Um deles, talvez o mais importante é o acesso a sites (e à descarga do seu conteúdo), passando pela análise e enviando/escrevendo no servidor, os dados obtidos assim como os resultados da análise.

Quando é feito o pedido ao servidor, este costuma demorar algum tempo a responder e a enviar. A tarefa é simples e os dados geralmente pequenos. É então possível usar o processamento dos clientes para fazer a mesma coisa, mas em vez de sequencial, fazê-lo em simultâneo, pedindo informações para vários sites ao mesmo tempo.

Pelos testes efectuados, esta possibilidade é real e efectiva. Foi possível usar em *Python* o módulo *threading (*para usar *multithreading* e o módulo *requests* para os pedidos de páginas web), para pedidos simultâneos poupando tempo. Na prática, no entanto, não se verificaram melhorias, bem pelo contrário, pois os pedidos eram poucos, e dado que o servidor web usa um cartão SD, a quantidade de escritas e de abertura de *threads* na base de dados poderia originar a tomada de sites em excesso para trabalhar, deixando outros nós sem "nada para fazer", potenciais corridas de *threads* que resultassem em erro e ficassem tarefas "penduradas", além da lentidão já observada com 2/3 equipamentos a enviar em simultâneo. Com o servidor web e a base de dados alojados num servidor "a sério", estes problemas de performance já não existem (*assim como as bases de dados não se corromperem com o volume de escrita" – isto verificou-se por duas vezes*).

5.2.2 Ferramentas do Servidor

A versão 1.0 tinha capacidades que a versão 2 (*ainda*) não tem, a nível de ferramentas secundárias. A nível de hardware um *Raspberry Pi* ainda não está capaz mas a nova arquitectura

versão 3 melhorou o desempenho. Os cartões SD são um problema e nota-se que à medida que o iKNOW vai aumentando a nível de base de dados, a resposta já começa a demorar pelo que um servidor padrão, poderá ser uma solução a avaliar para algo de produção.
Graficamente, o iKNOW pode ser melhorado.

Está previsto que o servidor não tenha qualquer interacção com o site-alvo. Assim, qualquer pedido na interface web, que actualmente esteja a fazer directamente (pedido de site por exemplo), seja inserido numa lista (*queue*) e seja depois processada, e enviada por um qualquer cliente. Isto impedirá que o site-alvo saiba quem fez a pesquisa, evitando assim que tenha conhecimento sequer do que é o iKNOW e impeça também o seu bloqueio.

5.2.3 Limites de memória

Alerta: na configuração actual, e como prova de conceito, não foram colocados limites ao que a página web/cliente pode pedir. Por isso, utilizando a página web "Gestão de OSINT's por categoria" para pedir todos as informações, rebentamos com o *script PHP* (derivado ao PHP estar a carregar toda a informação na memória).
Erro: "Fatal error: Allowed memory size of 134217728 bytes exhausted (tried to allocate 716800 bytes) in …"
A solução à partida é esvaziar os dados à medida que os for carregando, não os armazenando em memória. Outra, passa pela edição do ficheiro php.ini e inserindo "ini_set('memory_limit', '8192M');" no topo (mas é má prática pois nem todos os sistemas têm 8 gigas de RAM).
NOTA: O *Raspberry* pelos testes efectuados como servidor sofre um DoS se lhe pedirmos a disponibilidade dos equipamentos no iKNOW 1.0. Para DoS total, basta ir fazendo pedidos…

5.2.4 Comunicações seguras, SSH e Rede TOR

Na configuração actual, cliente e servidor podem fazer pesquisas na rede TOR. A configuração inicial corre um *script* que instala o cliente TOR, um serviço web e um acesso SSH. São opcionais o servidor SSH e a rede TOR.

Cada cliente pode ser contactado via rede TOR e SSH, independentemente de onde esteja. Isto possibilita/ria (se fosse o objectivo do trabalho), a criação de uma *botnet*. Cada cliente possui uma página web de nome thor.txt que indica para fins de *debug*, o IP interno, o IP externo, o seu nome(*hostname*) e o seu endereço *TOR* (.onion).

Pretende-se num futuro próximo, que os clientes usem a rede TOR para comunicarem dados para o servidor, para fazerem actualizações e para que fiquem protegidos na sua identidade, já que o ip da rede TOR não traduz a realidade da localização do seu utilizador.

5.2.5 Tempus fugit

Existem muitas funcionalidades que se perderam com a criação do novo iKNOW 2.0 e que são de muito interesse. Nomeadamente,

- a recuperação da interface das operações e suas tarefas;
- *dashboards* de disponibilidade, utilizadores, estatísticas de visitantes;
- oferta de largura de banda e processamento distribuído;

- granularidade do *logging* geral;
- entre outros que fazem com que o iKNOW 2.0 fique muito limitado à introdução de registos e à execução de relatórios;

Problemas como a escassez de tempo, emprego, mudanças de páginas web que estragavam o trabalho feito até então, mudanças nas aplicações que serviam de base, entre outras. Principalmente a saturação que chega quando o tempo foge...

5.3 Metodologia OSINT – Implementação com o iKNOW

Funcionamento sem metodologia OSINT: os dados em brutos de um motor de busca são aos milhares e pouco úteis. Obriga a que seja tomada atenção à hiperligação (origem dos dados é importante), assim como à data e aos termos localizados. Não é possível analisar toda a informação em tempo útil, especialmente se for em grande quantidade, correndo-se também o risco de perder informação.

Resultado: motor de busca devolve centenas de hiperligações que têm de ser percorridas manualmente com o risco de esquecermos algum e/ou de nos cansarmos antes de atingido o objectivo (localizar algo importante). Sem o recurso a motores de busca e/ou a sites/serviços do género, a procura de informação manual é ainda mais morosa, sujeita a erros e com maior probabilidade de não se encontrar o que se procura.

Funcionamento com OSINT: com uma abordagem metódica OSINT, as coisas podem ser bastante diferentes. Podemos automatizar o processo, criando e usando uma ferramenta que pesquise no motor de busca simulando um navegador *web/browser*. Os resultados obtidos podem ser filtrados para serem obtidas as hiperligações. Depois o simulador vai à página 2, 3... e repete o processo. Utilizando expressões regulares de forma constante em grandes quantidades de texto é possível varrer muito mais informação que manualmente, não querendo isto dizer que não se deva rever depois.

Resultado: temos agora imensa informação útil que pode ser armazenada para mais tarde ser processada, analisada por ferramentas analíticas, possibilitando criar alertas e descobrir, correlacionando dados, novas informações, gerando inteligência. Depois de criada a inteligência e seguindo o ciclo, segue-se o reporte da mesma, podendo ir acompanhado de outras informações que a suportem. A informação está estruturada. Ver a este respeito, a figura 8.

Um exemplo que não é OSINT mas faz esta correlação com resultados de diversas fontes, (às vezes até aparentemente desconexas) e nos fornece depois *intelligence* em tempo útil, são os SIEM[176], que dão um importante apoio aos analistas de segurança informática.

Após OSINT (ou outra qualquer ramificação de *intelligence*, desde que tratada com método e orientada ao que se pretende obter), temos informação útil que nos auxilia na decisão a tomar.

[176] SIEM - *Security Information and Event Management.* Exemplo: ArcSight, QRadar, FortiSIEM, Logrythm, AlienVault OSSIM, …

A imagem 54 pretende mostrar como, e o que se obtém usando OSINT. A técnica também é válida para outros tipos de informações.

Figura 54 - Depois de OSINT

5.4 Dificuldades, obstáculos e limitações técnicas do iKNOW

Existem limitações da ferramenta, não só técnicas, mas sobretudo devido à própria natureza do OSINT. É muito difícil identificar informação real de informação falsa por exemplo (tal como na vida real). Depois há outros factores que foram sendo identificados ou comprovados no decorrer da construção e utilização da ferramenta.

A construção iniciou-se em 2014 e no decorrer da mesma, que ainda hoje não está totalmente acabada (nem nunca estará já que para cada nova versão de sítio web ou de "protecção", é necessário adaptar com os ajustes devidos), verificou-se que os sites, ao longo do tempo, dificultaram a aquisição dos seus conteúdos ou utilizaram a *web* como arma de informação:

- *"Fake news"* - Blogues e sites web com notícias falsas ou patrocinados por empresas que colocam conteúdo ou palavras que podem comprometer a informação obtida automaticamente;
- Contra-informação ou contra-inteligência ciber – "Forças e Estados inimigos" podem disseminar em sites e redes sociais, informações falsas sobre um candidato (exemplo recente da ingerência russa nas eleições americanas) ou uma outra potência estrangeira, ou criar notícias sobre um potencial perigo para causar medo. As redes sociais são férteis em notícias falsas;
- Interesses económicos. Notícias falsas sobre economia, vendas, potencialidades de produtos podem alterar a percepção do consumidor e das notícias obtidas;
- Publicidade – muitos sites para ser alojados sem custos ou para ganhar dinheiro, são obrigados a aceitar conteúdos inseridos automaticamente por terceiros. O resultado é que um determinado site pode ser para crianças e a dada altura, ter publicidade para adultos. O site pode também conter links para sites pouco recomendados e que podem enganar ou "desviar a atenção da ferramenta".

Dificuldades encontradas:
A ferramenta cresceu e nem sempre foi para melhor, tendo sido feitos muitos avanços, mas possivelmente ainda mais retrocessos.

- Este projecto demorou anos desde o seu início até agora. A tecnologia avançou radicalmente e isto também se observa na evolução do iKNOW:
    - Python2 para Python3: código utilizado não funcionava em *python3*, devido à escassez de algumas bibliotecas e código deprecado;
    - PHP5 para PHP7: houve mudanças drásticas que destruíram a maior parte dos códigos, nomeadamente surgir de avisos e novos erros, com funções deprecadas;
- Pesquisas - Vários sites utilizam truques para aumentar o seu valor junto de motores de busca. Isto significa que irão aparecer primeiro no ranking e nos sites pesquisados do iKnow. Significa também que um dos truques é a utilização muitas vezes enganosa de palavras-chave;
- Publicidade e "cliques". Pior é a necessidade de agora muitos deles exigirem que o utilizador clique em botões "Clique para ver mais" ou em botões para fechar janelas que se abre sozinhas com publicidade antes de chegar ao conteúdo, que é o que realmente nos interessa;
- Muitos dos sites que realmente interessam, começaram a bloquear ou a dificultar as visitas de navegadores que não estejam identificados, ou que venham a partir da rede TOR, ou ainda que façam mais de x pedidos. Bloqueio de IP's. Alguns sites detectam muitos pedidos e bloqueiam o IP durante um tempo pré-definido por eles. Alguns sites da rede TOR, utilizam algumas manhas para detectar se é uma pessoa ou um automatismo a visitar o site. Como? Se um site visitar x *link* do próprio site, é adicionado à *firewall* e bloqueado a partir desse momento;
- Bloqueio de IP's via nacionalidade – alguns ips de origem estrangeira estão simplesmente bloqueados e é necessário recorrer a VPN para ultrapassar estas barreiras;
- Puzzles CAPTCHA. Quando alguns sites detectam que o ip de origem vem da rede TOR, é frequente a obrigatoriedade de introdução de um código de uma imagem, para se poder ver o conteúdo. Isto afecta claramente navegações em motores de busca e sites considerados "grandes";
- Decorrente do tempo e da constante evolução da ferramenta, houve grandes mudanças na programação das linguagens (nem sempre para melhor), das quais fica um exemplo: MyQSL e PHP. A mudança de PHP 5 para 7 estragou o código existente. Na questão do MySQL, onde antes se colocava todo o texto em utf-8, com "***mysql_set_charset('utf8', $link);".***

iKNOW – No decurso da escrita desta tese e da criação da ferramenta iKNOW, foi notória a transição entre o que inicialmente os sites de informação genérica forneciam e a cada vez maior colocação de entraves (imagens 55, 56 e 57) ao visitante e ao *crawler/scraper*, visto tão simplesmente, como a colocação de anúncios e a obrigatoriedade de clicar em botões para prosseguir para a informação que se pretende. Também por diversas vezes aconteceu que o site x onde se ia buscar informação era bloqueado judicialmente ou simplesmente saía do ar. Exemplos:

Figura 55 - Entraves colocados às ferramentas OSINT: publicidade e obrigatoriedade de clicar

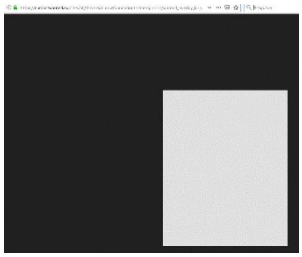


Figura 57 - Hiperligações falsas a proteger e a impedir a visualização de imagens

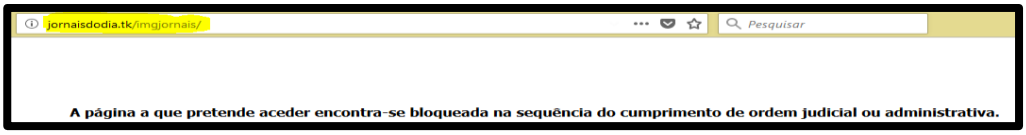


Figura 56 - Entraves à obtenção de OSINT: fontes desaparecem ou são bloqueadas

No caso da imagem 55, que é preciso clicar para ver, não é "grave", visto que era apenas um dos sítios onde a ferramenta iKnow ia buscar informação. Quando entramos na rede TOR, em busca de uma informação "importante" e uma hora ou duas depois, o site já não está acessível, o endereço TOR mudou ou somos bloqueados (a razão aqui pode ser variada: ip com muitos pedidos, ip do nó TOR de onde está a sair o pedido, bloqueio porque acedemos a um link que não era suposto por não ser visível na página, etc…). Aqui é pior pois a informação pode nunca mais ser vista.

Em progresso: O iknow pesquisa e faz *crawl/scrap* em sites. *onion* da rede *TOR*, no entanto recorre a *scripts* para o fazer. É utilizado um túnel que permite à ferramenta comunicar para aquela rede através de outros portos que não o porto 80 do protocolo *web*/HTTP. Está a ser melhorado para permitir a sua utilização de forma mais simples.

5.5 Testes, propostas e resultados dos inquéritos

Foram efectuados testes de usabilidade e satisfação relativamente à ferramenta de forma a auferir pontos positivos, pontos a melhorar e a experiência geral de utilização. Fica o resumo e os resultados.

5.5.1 Caracterização dos utilizadores que testaram a plataforma

Idade	20-30 anos	30-40 anos	40-50 anos	50-60 anos
Quantidade (8 pessoas)	1	4	2	1

Habilitações académicas	Licenciatura	Mestrado	Doutoramento	?
Quantidade (8 pessoas)	3	2	2	1

5.5.2 Primeiro teste - iKNOW versão 0.1

Formulário foi entregue depois do pedido de utilização livre do site. Foi pedido no decorrer da utilização que fosse pesquisado por um determinado tema à escolha do utilizador e depois por um tema no âmbito dos objectivos iKNOW.

Média de resultados

Grau de satisfação (1 a 5):	3
Facilidade de uso:	2
Utilidade (objectivos):	4
Experiência geral:	3
Dificuldades sentidas:	Passos para ligar ferramenta. Utilização. Utilidade.
Recomendações:	Aumentar a facilidade de uso. Obter páginas num formato que se pudesse usar. Gráficos.

5.5.3 Segundo teste - iKNOW versão 1.0

Formulário foi entregue depois do pedido de utilização livre do site. Foi pedido no decorrer da utilização que fosse pesquisado por um determinado tema à escolha do utilizador e depois por um tema no âmbito dos objectivos iKNOW.

Média de resultados

Grau de satisfação (1 a 5):	4
Facilidade de uso:	4
Utilidade (objectivos):	4
Experiência geral:	4
Dificuldades sentidas:	Alguma desorientação. Exportar informação.
Recomendações:	Novas capacidades melhoraram a experiência geral. Utilização sem problemas embora haja muitas opções e menus. Sugerido colocar logo ao inicio o que se pretende. Ideia dos *raspberrys* foi bem aceite tal como as comunicações, logins, etc.

5.5.4 Terceiro teste - iKNOW versão 2.0

Formulário foi entregue depois do pedido de utilização livre do site. Foi pedido no decorrer da utilização que fosse pesquisado por um determinado tema à escolha do utilizador e depois por um tema no âmbito dos objectivos iKNOW.

Média de resultados

Grau de satisfação (1 a 5):	5
Facilidade de uso:	5
Utilidade (objectivos):	5

Experiência geral:	4
Dificuldades sentidas:	Localizar conteúdos.
Recomendações:	Novas capacidades melhoraram a experiência geral. Utilização sem problemas. Foi apreciado ter logo o que se precisa no inicio. Gráficos também foram apreciados. Menus e aspecto do iKNOW 1.0 estavam melhores que estes embora o gráfico fosse bom.

A evolução dos resultados dos inquéritos foi positiva:

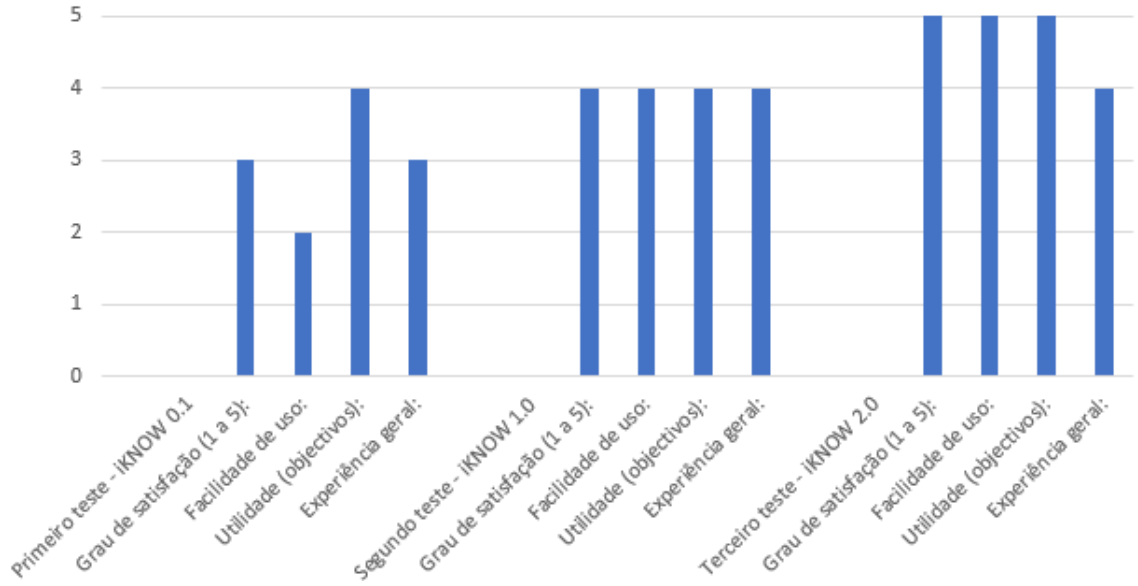


Figura 57 - Tabela de evolução de resultados dos inquéritos

5.5.5 Conclusões

Depois de definidos os objectivos para os sites escolhidos, pelas pessoas que testaram a ferramenta, foi unânime a vontade de ter algo mais simples nas operações, sem a tão (acreditava-se que fosse importante) necessidade de buscas contínuas de x em x tempo, mas sim, na obtenção directa da informação. Foi sugerida também a redução de "features" enquanto que outros acharam úteis, outros ainda acharam que era útil, mas sem necessidade de *login*.

Foram dadas diversas sugestões, abaixo listadas. Todas elas foram tomadas em conta e transpostas para uma nova versão, iKNOW 2.0, em que já foram implementadas.

Um dos pontos sugeridos, e que já haviam sido pensados, eram o de analisar o site, independente do seu conteúdo, e retirar apenas a notícia central. O texto do que realmente interessa, ignorando publicidades quer em texto quer em imagem, ignorando códigos *script*, ente outros. Foi sugerido até que o importante era o texto central. A técnica foi baseada na análise do site como um todo, verificação das zonas (*tags html*) e verificação de onde existe mais texto. Agora, com uma caixa de texto, basta inserir a url, clicar enviar, e numa nova página verificar se tal está de acordo com o que queremos ou não. A imagem 58 ilustra a simplicidade do pedido e/de aquisição de uma página web.

Obter noticia principal

Coloque a URL para capturar a principal noticia do site

URL

Enviar

Figura 58 - Introdução do endereço web pretendido

Introduzir | Sair sem fazer alterações

Introduzir nova informação

Edite e envie para actualizar registo.

Título | EUA lançam 'Sea Hunter', um navio drone que dispensa tripulação

Imagem

Figura 59 - Introdução do endereço web pretendido

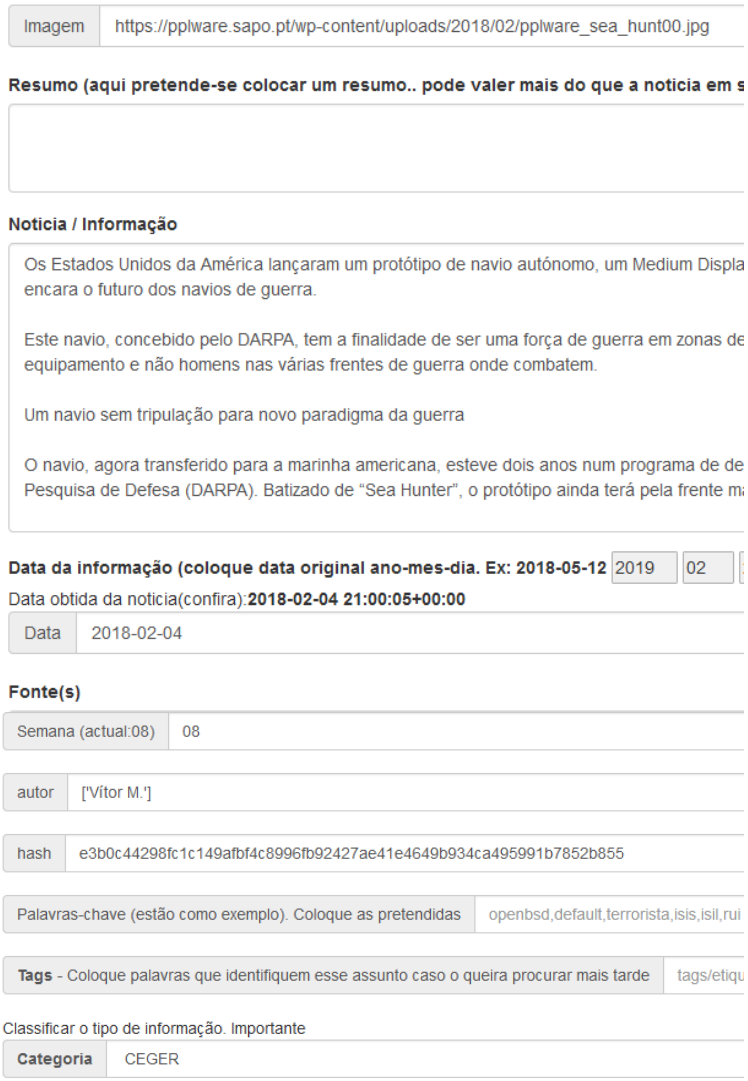


Figura 60 - Página mostra tudo o que foi recolhido. Pode ser alterado agora ou editado mais tarde.

Outros utilizadores sugeriram que o importante era criar uma "espécie de *feeds*" (imagem 61) de alguns sites, para consumo por outras ferramentas. Implementado! Em baixo, o exemplo dos tópicos retirados do site "thehackernews" (também foi implementado o mesmo para sites de emprego e sites de notícias). O iKNOW também consegue interpretar *feeds* que os sites possam ter.

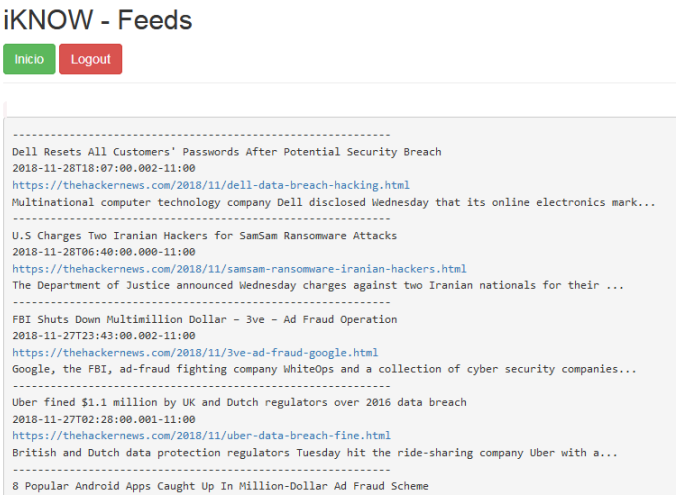


Figura 61 - Construção de relatórios tipo "feeds"

A nova versão também tem estatisticas dos OSINT inseridos, por categorias:

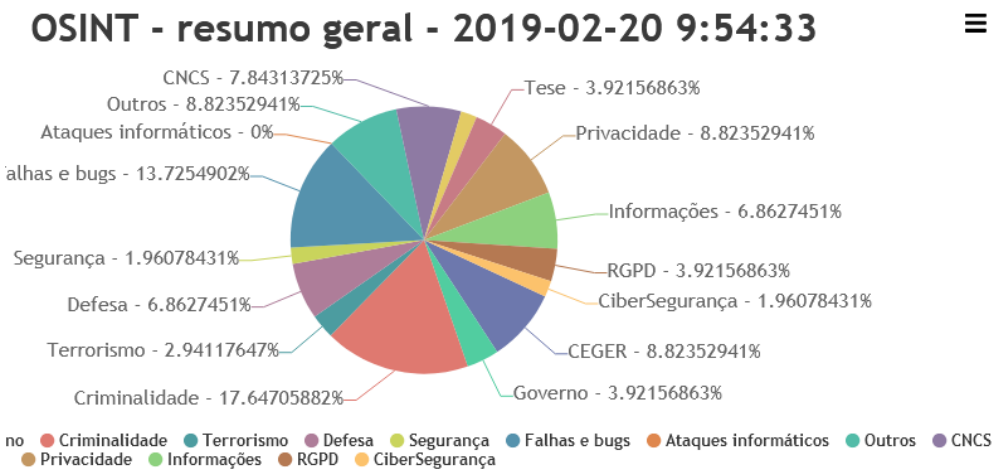


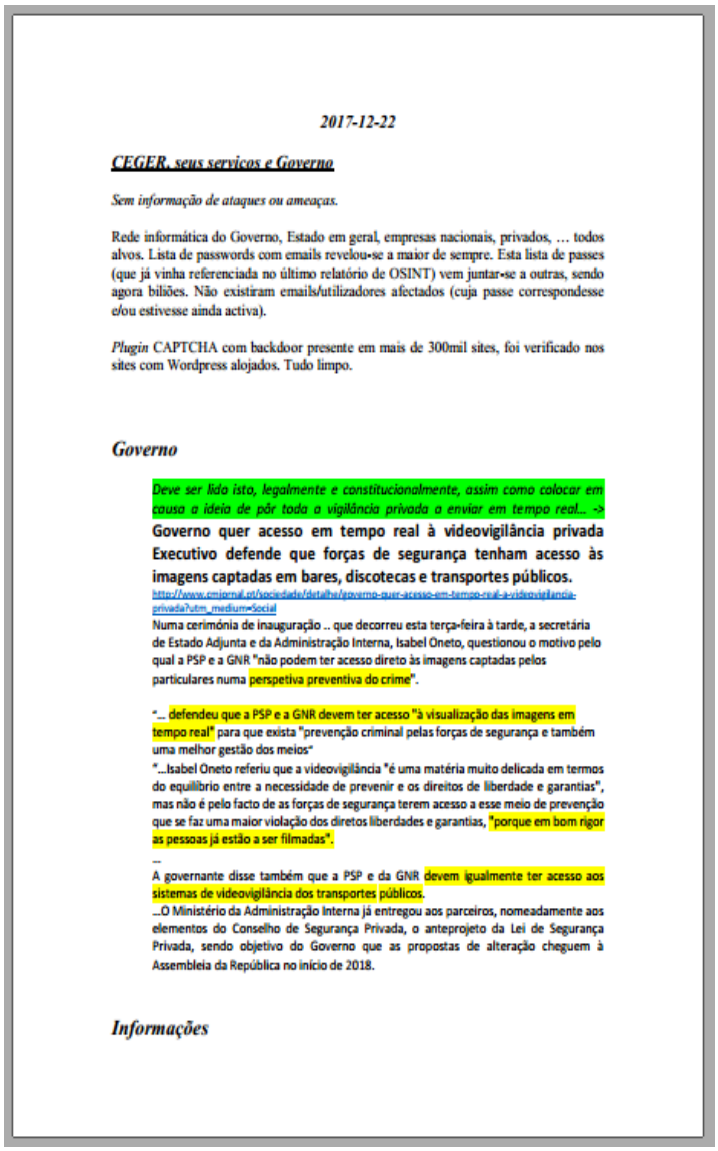
Figura 62 - Dashboards com os resultados OSINT obtidos, segundo quantidade e tema

5.5.1 Relatórios OSINT

No âmbito das tarefas diárias nos últimos locais onde o autor tem trabalhado, tem-se dedicado no conjunto das suas competências, à construção de relatórios, aos quais chamou "Relatórios OSINT" e que têm constituído um resumo semanal (*diário quando algo grave ou importante acontece*) do que vai acontecendo, com maior destaque ao local e ao objectivo da instituição onde se encontra.

Abaixo um exemplo do referido, destacando que a maior parte do mesmo já é automatizado pela ferramenta iKNOW. Estes relatórios seguem sempre com versão em texto no corpo do email e uma versão em PDF como anexo.

5.5.5.1 Exemplo de relatórios - iKNOW 1.0



2017-12-22

CEGER, seus serviços e Governo

Sem informação de ataques ou ameaças.

Rede informática do Governo, Estado em geral, empresas nacionais, privados, … todos alvos. Lista de passwords com emails revelou-se a maior de sempre. Esta lista de passes (que já vinha referenciada no último relatório de OSINT) vem juntar-se a outras, sendo agora biliões. Não existiram emails/utilizadores afectados (cuja passe correspondesse e/ou estivesse ainda activa).

Plugin CAPTCHA com backdoor presente em mais de 300mil sites, foi verificado nos sites com Wordpress alojados. Tudo limpo.

Governo

Deve ser lido isto, legalmente e constitucionalmente, assim como colocar em causa a ideia de pôr toda a vigilância privada a enviar em tempo real… ->

Governo quer acesso em tempo real à videovigilância privada
Executivo defende que forças de segurança tenham acesso às imagens captadas em bares, discotecas e transportes públicos.

http://www.cmjornal.pt/sociedade/detalhe/governo-quer-acesso-em-tempo-real-a-videovigilancia-privada?utm_medium=Social

Numa cerimónia de inauguração .. que decorreu esta terça-feira à tarde, a secretária de Estado Adjunta e da Administração Interna, Isabel Oneto, questionou o motivo pelo qual a PSP e a GNR "não podem ter acesso direto às imagens captadas pelos particulares numa perspetiva preventiva do crime".

"… defendeu que a PSP e a GNR devem ter acesso "à visualização das imagens em tempo real" para que exista "prevenção criminal pelas forças de segurança e também uma melhor gestão dos meios"

"…Isabel Oneto referiu que a videovigilância "é uma matéria muito delicada em termos do equilíbrio entre a necessidade de prevenir e os direitos de liberdade e garantias", mas não é pelo facto de as forças de segurança terem acesso a esse meio de prevenção que se faz uma maior violação dos diretos liberdades e garantias, "porque em bom rigor as pessoas já estão a ser filmadas".

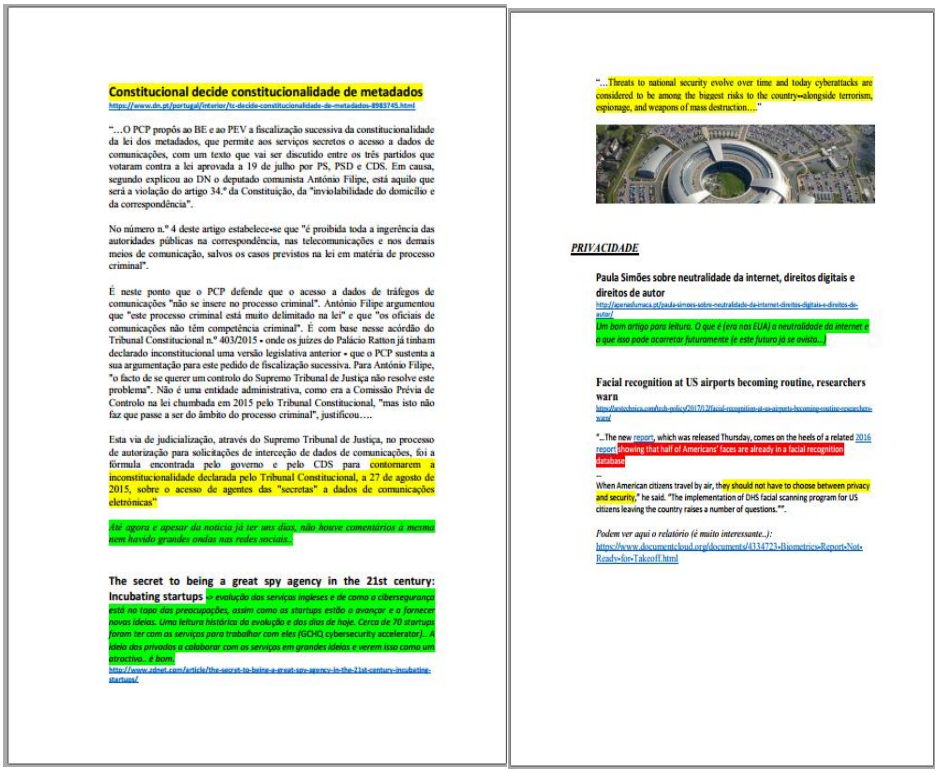
--

A governante disse também que a PSP e da GNR devem igualmente ter acesso aos sistemas de videovigilância dos transportes públicos.

…O Ministério da Administração Interna já entregou aos parceiros, nomeadamente aos elementos do Conselho de Segurança Privada, o anteprojeto da Lei de Segurança Privada, sendo objetivo do Governo que as propostas de alteração cheguem à Assembleia da República no início de 2018.

Informações

Figura 63 - Própria instituição e Governo



Constitucional decide constitucionalidade de metadados
https://www.dn.pt/portugal/interior/tc-decide-constitucionalidade-de-metadados-8983745.html

"…O PCP propôs ao BE e ao PEV a fiscalização sucessiva da constitucionalidade da lei dos metadados, que permite aos serviços secretos o acesso a dados de comunicações, com um texto que vai ser discutido entre os três partidos que votaram contra a lei aprovada a 19 de julho por PS, PSD e CDS. Em causa, segundo explicou ao DN o deputado comunista António Filipe, está aquilo que será a violação do artigo 34.º da Constituição, da "inviolabilidade do domicílio e da correspondência".

No número n.º 4 deste artigo estabelece-se que "é proibida toda a ingerência das autoridades públicas na correspondência, nas telecomunicações e nos demais meios de comunicação, salvos os casos previstos na lei em matéria de processo criminal".

É neste ponto que o PCP defende que o acesso a dados de tráfegos de comunicações "não se insere no processo criminal". António Filipe argumentou que "este processo criminal está muito delimitado na lei" e que "os oficiais de comunicações não têm competência criminal". É com base nesse acórdão do Tribunal Constitucional n.º 403/2015 - onde os juízes do Palácio Ratton já tinham declarado inconstitucional uma versão legislativa anterior - que o PCP sustenta a sua argumentação para este pedido de fiscalização sucessiva. Para António Filipe, "o facto de se querer um controlo do Supremo Tribunal de Justiça não resolve este problema". Não é uma entidade administrativa, como era a Comissão Prévia de Controlo na lei chumbada em 2015 pelo Tribunal Constitucional, "mas isto não faz que passe a ser do âmbito do processo criminal", justificou….

Esta via de judicialização, através do Supremo Tribunal de Justiça, no processo de autorização para solicitações de intercepção de dados de comunicações, foi a fórmula encontrada pelo governo e pelo CDS para contornarem a inconstitucionalidade declarada pelo Tribunal Constitucional, a 27 de agosto de 2015, sobre o acesso de agentes das "secretas" a dados de comunicações eletrónicas".

Até agora e apesar da noticia já ter uns dias, não houve comentários à mesma nem havido grandes ondas nas redes sociais..

The secret to being a great spy agency in the 21st century: Incubating startups *- evolução dos serviços ingleses e de como a cibersegurança está no topo das preocupações, assim como as startups estão a avançar e a fornecer novas ideias. Uma leitura histórica da evolução e dos dias de hoje. Cerca de 70 startups foram ter com os serviços para trabalhar com eles (GCHQ cybersecurity accelerator). A ideia dos privados a colaborar com os serviços em grandes ideias e verem isso como um atractivo.. é bom.*
http://www.zdnet.com/article/the-secret-to-being-a-great-spy-agency-in-the-21st-century-incubating-startups/

"…Threats to national security evolve over time and today cyberattacks are considered to be among the biggest risks to the country—alongside terrorism, espionage, and weapons of mass destruction…."

PRIVACIDADE

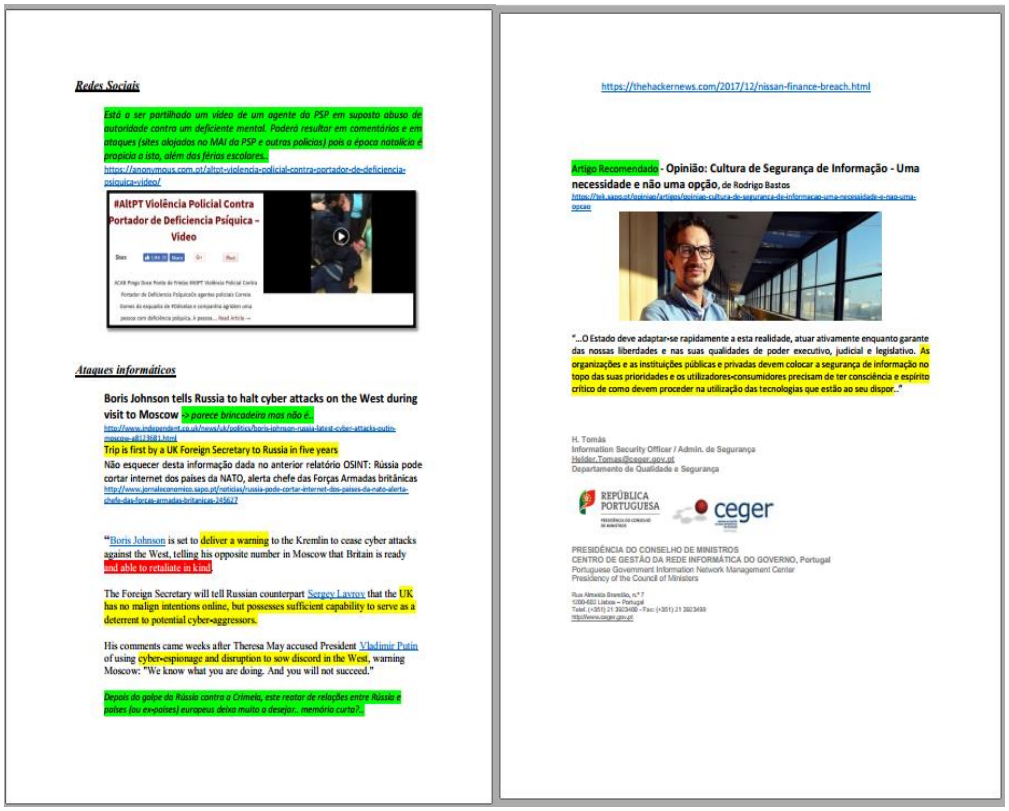
Paula Simões sobre neutralidade da internet, direitos digitais e direitos de autor
http://apeniasfumaca.pt/paula-simoes-sobre-neutralidade-da-internet-direitos-digitais-e-direitos-de-autor/
Um bom artigo para leitura. O que é (era nos EUA) a neutralidade da internet e o que isso pode acarretar futuramente (e este futuro já se avista…)

Facial recognition at US airports becoming routine, researchers warn
http://arstechnica.com/tech-policy/2017/07/facial-recognition-at-us-airports-becoming-routine-researchers-warn/

"…The new report, which was released Thursday, comes on the heels of a related 2016 report showing that half of Americans' faces are already in a facial recognition database.
…
When American citizens travel by air, they should not have to choose between privacy and security," he said. "The implementation of DHS facial scanning program for US citizens leaving the country raises a number of questions.*"*.

Podem ver aqui o relatório (é muito interessante..):
https://www.documentcloud.org/documents/4134723-BiometricsReport-Not-Ready-for-Takeoff.html

Figura 64 - Informações e Privacidade



Redes Sociais

Está a ser partilhado um vídeo de um agente da PSP em suposto abuso de autoridade contra um deficiente mental. Poderá resultar em comentários e em ataques (sites alojados no MAI da PSP e outros policias) pois a época natalícia é propicia a isto, além das férias escolares..
https://anonymous.com.pt/altpt-violencia-policial-contra-portador-de-deficiencia-psiquica-video/

Ataques informáticos

Boris Johnson tells Russia to halt cyber attacks on the West during visit to Moscow *- parece brincadeira mas não é..*
http://www.independent.co.uk/news/uk/politics/boris-johnson-russia-latest-cyber-attacks-putin-moscow-a8123981.html
Trip is first by a UK Foreign Secretary to Russia in five years
Não esquecer desta informação dada no anterior relatório OSINT: Rússia pode cortar internet dos países da NATO, alerta chefe das Forças Armadas britânicas
http://www.jornaleconomico.sapo.pt/noticias/russia-pode-cortar-internet-dos-paises-da-nato-alerta-chefe-das-forcas-armadas-britanicas-245627

"Boris Johnson is set to deliver a warning to the Kremlin to cease cyber attacks against the West, telling his opposite number in Moscow that Britain is ready and able to retaliate in kind.

The Foreign Secretary will tell Russian counterpart Sergey Lavrov that the UK has no malign intentions online, but possesses sufficient capability to serve as a deterrent to potential cyber-aggressors.

His comments came weeks after Theresa May accused President Vladimir Putin of using cyber-espionage and disruption to sow discord in the West, warning Moscow: "We know what you are doing. And you will not succeed."

Depois do golpe da Rússia contra a Crimeia, este reator de relações entre Rússia e países (ou ex-países) europeus deixa muito a desejar.. memória curta?..

https://thehackernews.com/2017/12/nissan-finance-breach.html

Artigo Recomendado - **Opinião: Cultura de Segurança de Informação - Uma necessidade e não uma opção,** de Rodrigo Bastos
https://twit.sapo.pt/opiniao/artigos/opiniao-cultura-de-seguranca-de-informacao-uma-necessidade-e-nao-uma-opcao

"…O Estado deve adaptar-se rapidamente a esta realidade, atuar ativamente enquanto garante das nossas liberdades e nas suas qualidades de poder executivo, judicial e legislativo. As organizações e as instituições públicas e privadas devem colocar a segurança de informação no topo das suas prioridades e os utilizadores-consumidores precisam de ter consciência e espírito crítico de como devem proceder na utilização das tecnologias que estão ao seu dispor.."

H. Tomás
Information Security Officer / Admin. de Segurança
Helder.Tomas@cegger.gov.pt
Departamento de Qualidade e Segurança

REPÚBLICA PORTUGUESA
ceger

PRESIDÊNCIA DO CONSELHO DE MINISTROS
CENTRO DE GESTÃO DA REDE INFORMÁTICA DO GOVERNO, Portugal
Portuguese Government Information Network Management Center
Presidency of the Council of Ministers

Rua Almeida Brandão, n.º 7
1200-620 Lisboa – Portugal
Telef. (+351) 21 3923400 - Fax: (+351) 21 3923499
http://www.cegger.gov.pt

Figura 65 - Redes sociais e ataques informáticos

Do que se obtém automaticamente, é necessário sempre rever. Maior parte das vezes a simples descoberta de um ou mais termos, não tem associada uma importância de relevo, quer por ser antigo, quer por ser um termo colocado no site apenas para ser encontrada nos motores de busca (truques para subir no *ranking* e optimizações de SEO[177]). A necessidade de revisão manual é obrigatória para assegurar também que uma informação pode não ser útil para entidade A ou B, mas ser para entidade C. A revisão apenas é possível se o volume de informação não for absurdamente alto, perdendo-se aí na tradução, tempo útil, às vezes escasso para evitar uma situação.

O sistema de relatórios iKNOW já está em produção no local de serviço do autor.

5.5.5.2 Exemplo de relatórios - iKNOW 2.0

Na figura 66 (lado esquerdo) temos o exemplo de um relatório PDF, gerado para a instituição (*simbolo propositadamente e manualmente apagado*). Contém a identificação da instituição, título, semana, data, um pequeno gráfico e, agrupado por categorias, as noticias.
Todas as semanas em que foi gerado relatório, ficam a fazer parte de uma base de dados acessível dentro da ferramenta (*figura 66 – lado direito*)

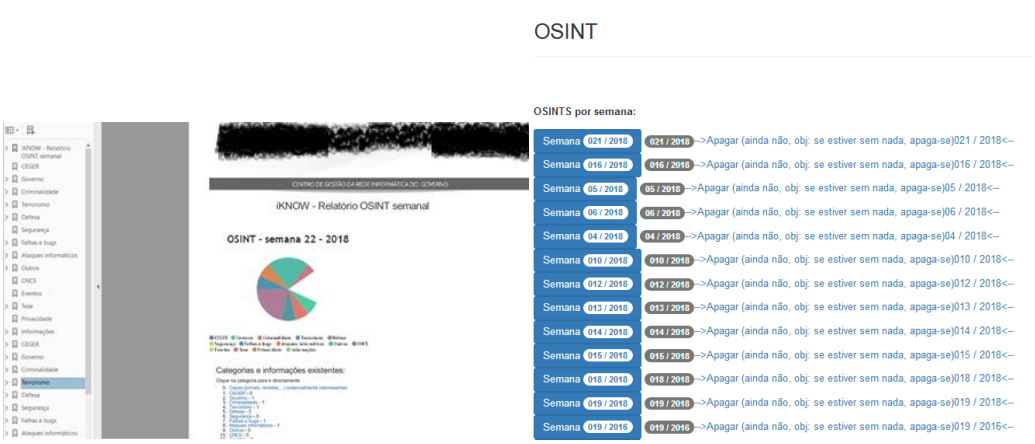
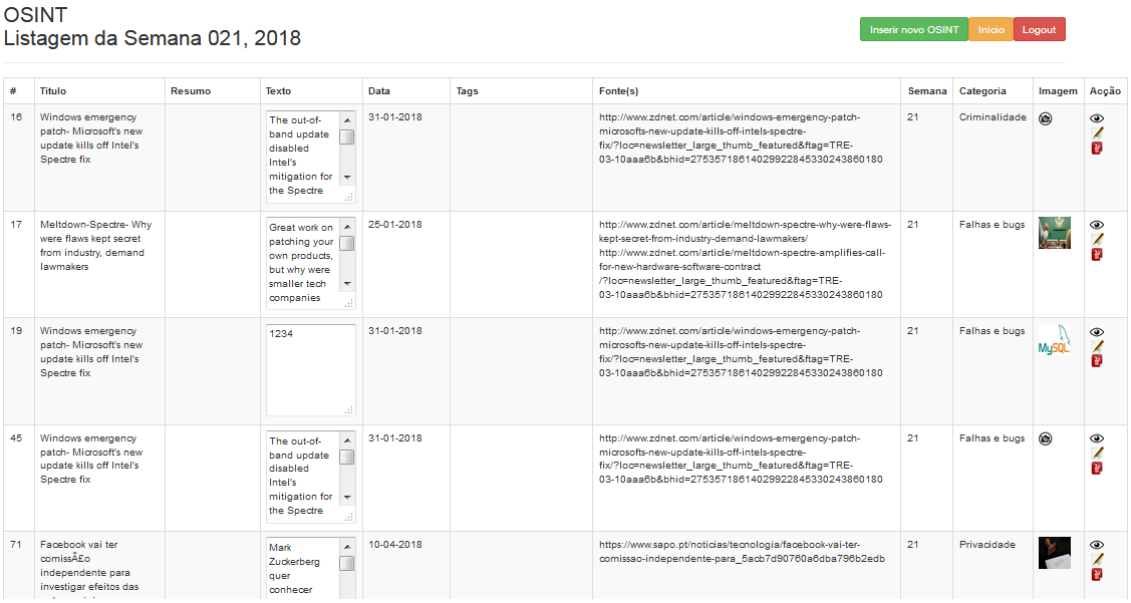


Figura 66 – iKNOW 2.0 Relatório Semanal (instituição propor)

[177] SEO – Search engine optimizations

Antes de ser gerado o relatório final, é possível verificar e confirmar todos os itens (imagem 67).


OSINT
Listagem da Semana 021, 2018

Inserir novo OSINT | Inicio | Logout

#	Titulo	Resumo	Texto	Data	Tags	Fonte(s)	Semana	Categoria	Imagem	Acção
16	Windows emergency patch- Microsoft's new update kills off Intel's Spectre fix		The out-of-band update disabled Intel's mitigation for the Spectre	31-01-2018		http://www.zdnet.com/article/windows-emergency-patch-microsofts-new-update-kills-off-intels-spectre-fix/?loc=newsletter_large_thumb_featured&ftag=TRE-03-10aaa6b&bhid=27535718614029922845330243860180	21	Criminalidade		
17	Meltdown-Spectre- Why were flaws kept secret from industry, demand lawmakers		Great work on patching your own products, but why were smaller tech companies	25-01-2018		http://www.zdnet.com/article/meltdown-spectre-why-were-flaws-kept-secret-from-industry-demand-lawmakers/ http://www.zdnet.com/article/meltdown-spectre-amplifies-call-for-new-hardware-software-contract /?loc=newsletter_large_thumb_featured&ftag=TRE-03-10aaa6b&bhid=27535718614029922845330243860180	21	Falhas e bugs		
19	Windows emergency patch- Microsoft's new update kills off Intel's Spectre fix		1234	31-01-2018		http://www.zdnet.com/article/windows-emergency-patch-microsofts-new-update-kills-off-intels-spectre-fix/?loc=newsletter_large_thumb_featured&ftag=TRE-03-10aaa6b&bhid=27535718614029922845330243860180	21	Falhas e bugs		
45	Windows emergency patch- Microsoft's new update kills off Intel's Spectre fix		The out-of-band update disabled Intel's mitigation for the Spectre	31-01-2018		http://www.zdnet.com/article/windows-emergency-patch-microsofts-new-update-kills-off-intels-spectre-fix/?loc=newsletter_large_thumb_featured&ftag=TRE-03-10aaa6b&bhid=27535718614029922845330243860180	21	Falhas e bugs		
71	Facebook vai ter comissÃ£o independente para investigar efeitos das		Mark Zuckerberg quer conhecer	10-04-2018		https://www.sapo.pt/noticias/tecnologia/facebook-vai-ter-comissao-independente-para_5acb7d90760a6dba796b2edb	21	Privacidade		

Figura 67 - itens para relatório semanal iKNOW 2.0

Também é possível gerar relatórios com capas de jornais ou outras imagens(figura 68):


OSINT - jornais

Inserir | Inicio | Logout

Capas de jornais por semana:

Semana	Acção	Gerar relatório	Teste relatório semanal
39	Ver tudo de semana 39,2018	Relatório 39,2018	Gerar semana 39,2018
45	Ver tudo de semana 45,2018	Relatório 45,2018	Gerar semana 45,2018
48	Ver tudo de semana 48,2018	Relatório 48,2018	Gerar semana 48,2018
21	Ver tudo de semana 21,2019	Relatório 21,2019	Gerar semana 21,2019
22	Ver tudo de semana 22,2019	Relatório 22,2019	Gerar semana 22,2019
28	Ver tudo de semana 28,2019	Relatório 28,2019	Gerar semana 28,2019
29	Ver tudo de semana 29,2019	Relatório 29,2019	Gerar semana 29,2019

Todas as capas:

Data	Semana	Dia	Mês	Ano	Acção	Acção 2
29092018	39	29	09	2018	29-09-2018	Apagar (inactivo)
11112018	45	11	11	2018	11-11-2018	Apagar (inactivo)
29112018	48	29	11	2018	29-11-2018	Apagar (inactivo)

Figura 68 - iKNOW 2.0 - relatórios – jornais

5.5.5.3 Exemplo de busca de termos – *Fuzzy string matching*

A pesquisa de termos é um dos objectivos desta dissertação e projecto, pois é com ele que podemos pesquisar material roubado, localizar pessoas, entre outros. Um dos objectivos é diariamente fazer pesquisas de material roubado em sites de vendas de produtos em segunda mão.

A figura 69 mostra o resultado de uma busca por termos. Existe uma lista de termos que pode estar em base de dados ou ficheiro. Com o crawler anterior, era possível varrer nas páginas

pesquisadas, uma busca por estes termos, com recurso a *fuzzing*[178]. Quantos mais termos existissem, maior seria o resultado depois em % de eficácia e pontuação. Está a ser implementado no iknow 2.0, e já existe a página a explicar funcionamento técnico.
O funcionamento é feito com a introdução da operação, ou, no estado actual, do nome de quem pede. Ao clicar pesquisar, é feita uma chamada a um *script* que vai executar diversos comandos tais como pedir o site, pesquisar a informação, calcular pontuação e gerar ficheiro PDF.

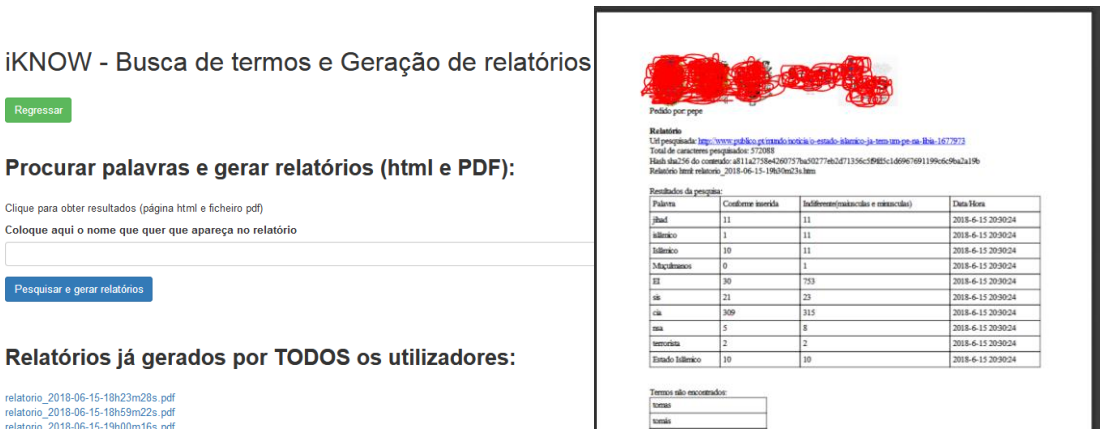


Figura 69 - Pesquisa por termos e resultado

Depois de tudo feito, é inserida na base de dados, a referência ao novo ficheiro criado. A imagem 70 mostra o funcionamento do *script* da linha de comandos. Podemos ver as palavras da lista, se foram encontradas e se sim, se em minusculas ou maiusculas. Esta diferenciação é importante porque podemos encontrar "ei" cinquenta vezes e é um "olá". Se encontrarmos "EI"em maiusculas, podemos estar perante uma página do Estado Islâmico. A técnica de *fuzzing* é feita recorrendo a um algoritmo que ignora caracteres que possam existir na palavra, propositadamente ou não. Exemplo para secreto: s3cr3t0, secret0, xecreto, secret0, s..e..c.re..to, … O resultado para cada palavra é dado numa percentagem de semelhança. Foi utilizada para este efeito, a biblioteca Python, *Fuzzywuzzy*[179].

Em baixo, na imagem 70 vemos a execução da pesquisa a uma lista de sitios web, palavras encontradas e geração do relatório html (é feito no momento e automaticamente, uma página web com imagens previamente escolhidas por nós, e com código html pré-feito que depois é completado com os novos valores e transformado em PDF de seguida).

[178] *Fuzzing* – no sentido geral, é uma técnica de geração aleatória de caracteres e valores. Neste caso, pretende-se que sejam obtidas palavras, mesmo que pelo meio existam alguns caracteres para enganar analisadores automáticos de texto.
[179] *fuzzywuzzy* – biblioteca Python para fuzzy string matching, disponível online em https://github.com/seatgeek/fuzzywuzzy

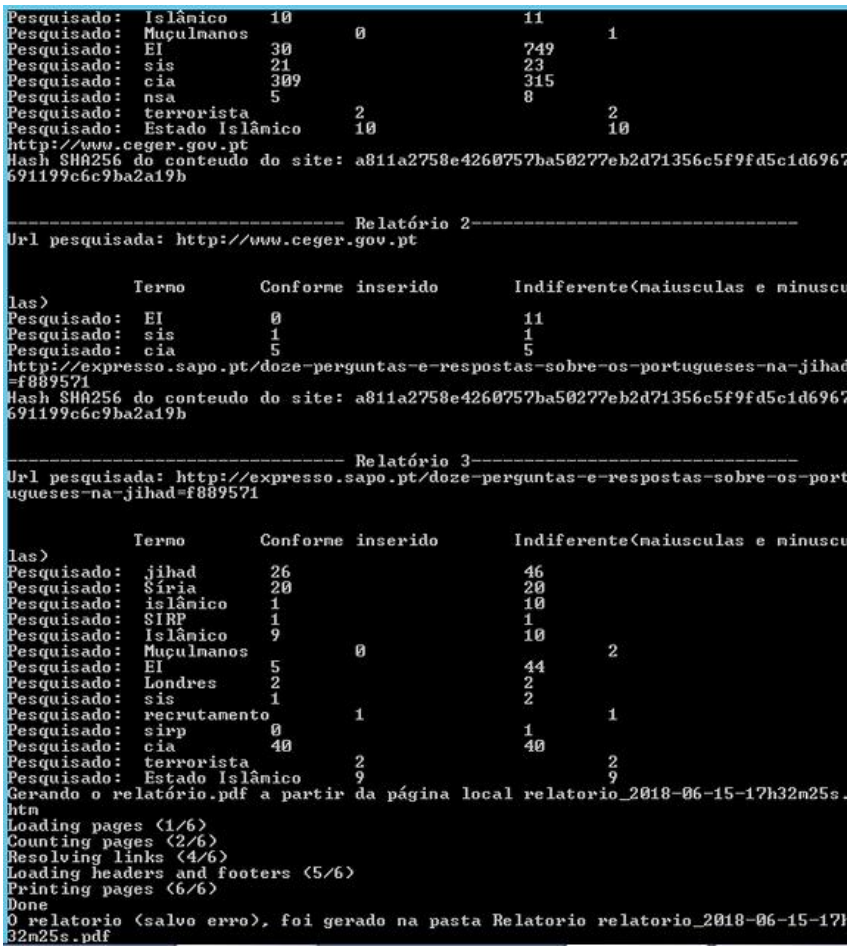


Figura 70 - funcionamento do scrapper de pesquisa de termos

CAPÍTULO VI - Conclusões

6.1 OSINT - A crescente necessidade e importância

Sabendo nós que hoje em dia praticamente tudo e todos estão na Internet e nas fontes abertas, sabendo nós também que a Internet é o meio de acesso e partilha de informação mais universal na actualidade (e possivelmente desde sempre). Dizendo a própria CIA que a informação não precisa ser secreta para ser de extrema importância…

Então, deve ser utilizado o OSINT (sem desdenhar dos outros tipos de inteligência, porque cada um tem o seu papel e não se pode simplesmente abandonar as outras) como uma abordagem holística para resolver problemas, de uma forma menos dependente de motores de busca. No fundo, ser o utilizador a resolver o seu próprio problema, capacitando-o para não depender dos motores de busca, mas sim nos seus próprios meios de análise e ferramentas.

Um *paper*[180] da Kapow Katalyst, tem um texto na sua ferramenta que vem de encontro ao que se pretende: *"OSINT data sources are as varied as the Internet itself. Mission-critical data can reside in blogs, in news feeds, in social media-and can even be hosted on short-lived sites reside in blogs, in news feeeds, in on the dark web. As technology standards continue to evolve one thing is certain: OSINT data sources will remain a moving target-in more ways than one."*
Porquê OSINT:

- Vivemos num mundo que precisamos da informação rapidamente
- Precisamos da informação agora
- Há decisões a esperar resposta
- Tudo e todos, queiram ou não, já estão na internet
- Auxílio agora, rápido e eficaz a forças policiais e de investigação
- As fotos também deixam rasto
- Testar o anonimato, vendo "como os outros (nos) vêm"
- Descobrir, evitar ou reduzir prejuízos de eventos (ataques informáticos, ataques a reputação, esquemas de *phishing*, …) antes de acontecerem
- Co-relacionamento e detecção de insiders, sistemas infectados, funcionários descontentes, … (atenção que dependendo da política que deve existir no local de trabalho/organização, podemos estar a transgredir as novas Leis de Privacidade e Dados pessoais):
    - O que faz o funcionário a quem foram detectadas comunicações para domínios maliciosos? Porque razão usa portos especiais? Tem as permissões correctas ás suas funções? …
    - Possibilidade de fazer buscar na *deep web*?

O OSINT não permite um trabalho mais fácil, nem mais rápido nem sequer mais fiável que outras fontes (HUMINT e espionagem por exemplo), já que, como veremos, existe demasiada informação de pouca qualidade e muito dispersa por muitas fontes, o que obriga a uma melhor selecção e do que realmente interessa. Por outro lado, não implica informações que possam colocar problemas (problemas com a Lei, direitos de autor, etc) mais tarde.

[180] *Paper* disponível na web em https://wikileaks.org/spyfiles/docs/KAPOWSOFTWARE-2011-BuilyourOSIN-en.pdf, acedido em 2017-12-05

"The challenge is being able to find the strategic information hidden massive volumes of data available on the web and on public sources.[181]*"*

OSINT não é necessariamente obtido usando *software Open Source* nem estes significam a mesma coisa. *Software Open Source* segundo a *wikipedia*[182], é software cujo código-fonte é disponibilizado de forma a ser estudado, alterado e melhorado. Existem vários tipos de licença e utilizações, mas deixarei isso para o leitor, embora o recomende (*e veja neste tipo de software grandes economias de tempo, recursos humanos e uma possibilidade infinita de entreajuda e evolução do software, sem se inventar a roda permanentemente*). Exemplo prático: o *software* open-source *NMAP* permite-nos descobrir muitas informações sobre uma determinada rede e serviços, mas este tipo de informações não são à partida públicos, e as informações recolhidas por este *software* podem ser consideradas já espionagem e em certos estados/países, é considerado pela Lei como crime (*citando um amigo, é como experimentar fechaduras de automóveis "só" para ver se abre..).*

Pesquisando diversas fontes sobre um mesmo assunto, podemos descobrir informações muito diversas sobre indivíduos, empresas ou outros. Por exemplo, a partir do nome, podemos descobrir onde trabalha (*Linked.In*), gostos e amigos (*FaceBook*), localização geográfica, (*Facebook* com *GoogleMaps* e/ou *Twitter*), entre tantas outras. Podemos depois cruzar essa informação e assegurar-nos que o que descobrimos é fiável (já que um dos principais problemas da Internet e do OSINT é a quantidade muito grande de informação que surge numa primeira abordagem).

Actualmente começa a ser um padrão, o uso de OSINT nas empresas para recrutamento[183] (*empresa verifica nas redes sociais se candidato falou verdade, se aparenta entrar no "perfil" pretendido,…),* pelo Governo para averiguação de riqueza e taxação, pelo cidadão comum, para pesquisar mais sobre a pessoa, o seu trabalho e os seus gostos pessoais, entre outros. Também nas empresas se assiste ao cada vez maior uso de OSINT naquilo a que hoje podemos chamar de *Competitive Intelligence*. A Internet veio possibilitar uma maior capacidade de OSINT (além de uma quase total falta de privacidade).

A história mostrou como o uso do OSINT permitiu antever alguns acontecimentos, e como foi por diversas vezes ignorado. Os usos do OSINT com outras *inteligências* poderiam ter evitado alguns ataques?

O serviço FIBS (ver evolução histórica) foi criado após o ataque sofrido de Pearl Harbor. Mais tarde, e devido ao ataque das Torres Gémeas[184], o presidente foi aconselhado a dar mais importância à OSINT e surgiu o Centro de OSINT, OSC(Open Source Center).

Segundo um *site* web (*expertsystem*[185]) dedicado a informações, "***OSINT plays a key role in national security.*** *But the value of Open Source Intelligence has not been limited to the military*

[181]http://www.expertsystem.com/value-open-source-intelligence-21st-century/

[182]https://en.*wikipedia*.org/wiki/Open-source_software

[183] Sendo o LinkedIn(https://www.linkedin.com/), um dos maiores sites da actualidade na área da procura de emprego e recrutamento

[184] Ataque terrorista, recorrendo a 2 aviões, que causou a queda de dois edificios simbólicos dos EUA, a 11 de Setembro de 2001

[185]http://www.expertsystem.com/value-open-source-intelligence-21st-century/

context: OSINT is ***equally helpful in collecting the strategic information that allows companies and businesses to empower the decision-making process***"**.**

Importância cada vez maior do OSINT:

- Existem muitas fontes importantes de informações não classificadas, gratuitas e disponíveis
- As redes sociais e a Internet fervilham de informação, de forma contínua e em tempo real (redes sociais *Facebook's*, redes de emprego *LinkedIn*, sites de encontros, fóruns, sites temáticos, …)
- Com tempo e recursos humanos, toda esta informação pode ser analisada, filtrada e gerados conhecimentos úteis (*se não para agora, se calhar para amanhã quando quisermos perceber o porquê de um ataque informático, cibernético ou armado contra os nossos*).
- Uma pesquisa rápida na Internet permitiu observar uma grande quantidade de livros, sítios web, cursos superiores, formações (entre as quais, até uma formação[186] leccionada à Polícia de Segurança Pública sobre OSINT
- "*A recolha de informação criminal em fontes abertas constitui uma importante ferramenta policial*".[187]

6.2 Comparativo OSINT vs HUMINT

A referência ao tema de HUMINT, nesta dissertação e projecto sobre OSINT, pretendeu informar e comparar OSINT e HUMINT nas suas diversas formas (a espionagem, agentes encobertos, agentes infiltrados e *insiders*), assim como as suas ambições, técnicas e protagonistas, e a sua relação com OSINT. Poderemos falar de rivalidade OSINT vs HUMINT?

Se no passado havia menos fontes de informação e a HUMINT era a única forma de obter informação concreta e de valor, a tempo, hoje, isto parece ter mudado.

É opinião do autor desta dissertação que com o actual rumo de acontecimentos que se vem vindo a assistir, com a maior quebra de privacidade e com a crescente despreocupação das camadas etárias mais novas, que futuramente, todas as informações estarão na rede, mais ou menos acessíveis. Portanto irá haver ainda mais OSINT. Também com o evoluir dos sistemas actuais de informação, haverá maior integração de sistemas (o que era será(?) positivos para as Policia e serviços de informações) e futuramente teremos um ainda maior *Big Brother* que terá acesso a todos os sistemas (COMINT) e informações, relevando para segundo lugar[188], a necessidade de HUMINT. Independentemente disto, a ameaça de pessoal interno é a maior ameaça para qualquer organização/Governo e mereceu uma atenção especial.

Hoje em dia, temos disponíveis para todos, algo que antigamente era muito difícil de obter como por exemplo, a localização de embarcações marítimas[189], aviões (no site

[186] "A CyberS3c.pt esteve hoje na Polícia de Segurança Pública, a dar um curso de OSINT, recolha e tratamento de dados em fontes abertas, foi extremamente recompensador ter uma plateia tão atenta e participativa." – Sérgio Silva, LinkedIN, 22-02-2019

[187] O conceito de Fontes Abertas na Investigação do Ciber Crime, 2014, disponível *online* em http://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime

[188] Em termos quantitativos de recursos e quantidade, não em termos de qualidade

[189] https://www.marinetraffic.com/

FlightRadar[190] até são facultadas instruções para ajudarmos com equipamentos), pessoas. Tudo isto está disponível na Internet. OSINT puro. Não há necessidade de HUMINT, correcto? Não. O HUMINT saberá se o avião x vai realmente levantar vôo e o que tem a bordo. Mais.. no caso de uma guerra ou de um simples mudar de ideias, estas informações públicas poderão desaparecer, assim como, ninguém nos garante que TODOS os aviões estão ali listados (militares?)..

Considera-se que é fundamental saber operar nas fontes abertas, podendo quiçá, evitar a perda de vidas humanas em actos HUMINT, preparando atempadamente a ida dos elementos HUMINT até aos seus alvos, recolhendo provas para gerar objectivos, recolhendo opiniões dos povos nativos antes de entrarem os militares, recolhendo mapas das instalações, horários de guardas, entre outros. Tudo isto também é válido para ataques de *pentesting* físico.

Em termos de credibilidade, a OSINT pode ser mais facilmente defendida, afinal é informação acessível a todos e não existem *à priori*, problemas por apresentar uma informação que todos sabem.

A informação de qualidade como diz a própria CIA[191], não precisa de ser informação classificada para ser de elevada importância: "***Information does not have to be secret to be valuable****. Whether in the blogs we browse, the broadcasts we watch, or the specialized journals we read, there is an endless supply of information that contributes to our understanding of the world."*

Já com fontes fechadas o problema é exactamente o oposto "…*os espiões debatem-se com a credibilidade. Para obter segredos, os espiões têm de ser traidores. Têm de trair o seu país e contar mentiras àqueles que estão à sua volta…É difícil ter a certeza que uns mentirosos tão rotinados e talentosos não mentem também acerca da informação que fornecem*"[192]
Seja como for, só a utilização de agentes (HUMINT) no terreno, permite averiguar com total certeza se os media daquele país, ou se a informação propagandeada, corresponde à verdade. Para um analista de fora, OSINT são as informações públicas e estas nem sempre retratam a verdade. Tal como só a utilização de agentes infiltrados permite saber a extensão e verdadeira capacidade.

6.2 O projecto iKNOW em números e imagens

"Geralmente, não há diferença entre administrar muitos e administrar poucos.
Trata-se de uma questão de organização."
Sun Tzu,em"A Arte da Guerra"

O projecto demorou anos a desenrolar-se e por isso foi crescendo em tamanho e complexidade. Algumas ideias viram a luz do dia, outras "fundiram". Ficam alguns números e imagens da plataforma e materiais utilizados:

- 45 equipamentos *Raspberry* de diversos modelos;

[190] https://www.flightradar24.com/38.76,-9.02/12
[191] Disponível online em https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html, consultado 28/08/2018
[192] Agentes secretos, a nova espionagem mundial. pp 65.

- 50 cartões SD de diversas marcas e tamanhos;
- 40 cabos eth 5e, 14 cabos eth 6;
- 3 routers wireless Asus, 2 routers wireless Tp-Link;
- 2 *routers* cisco empresariais, 2 *smart switches* Tp-Link, 2 *switches* Tp-Link;
- 2 relés, 4 *breadboards*, 8 extensões, 2 monitores, 3 bolsas de transporte, 3 mini-switches D-Link, 10 carregadores USB, 50 cabos usb, …

Algumas imagens:

Figura 71 - iKNOW em números e imagens – 50 cartões e 11 equipamentos de rede

Figura 72 - iKNOW – 3 bolsas de transporte e alguns equipamentos

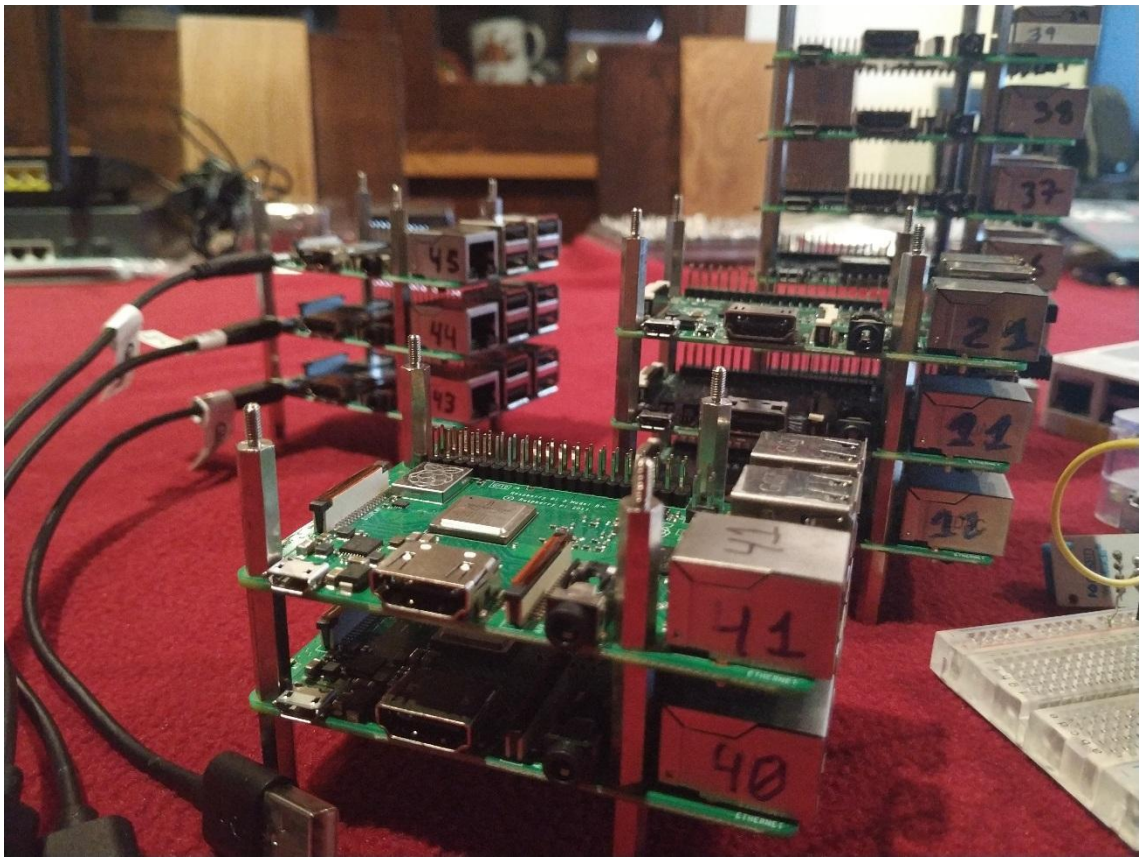


Figura 73 - o exército e plataforma iKNOW em mudanças

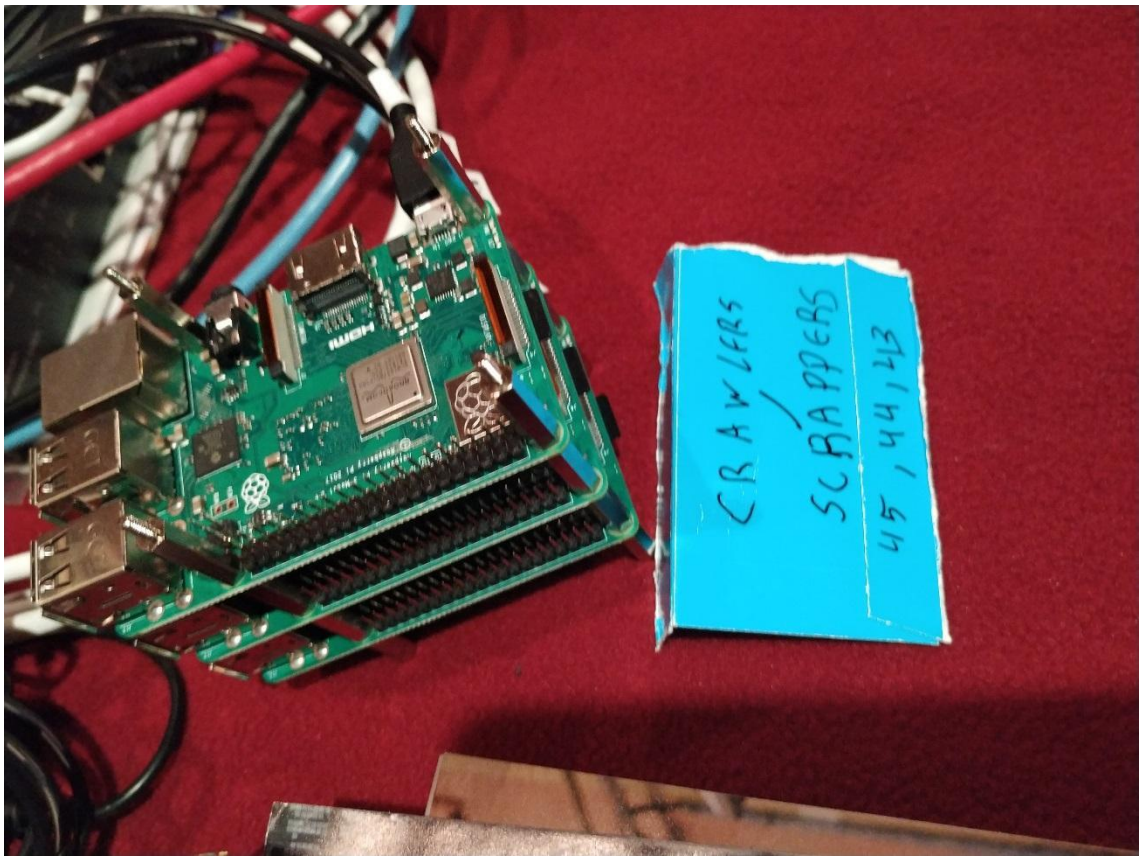


Figura 74 - Cluster de 3 crawlers / scrappers

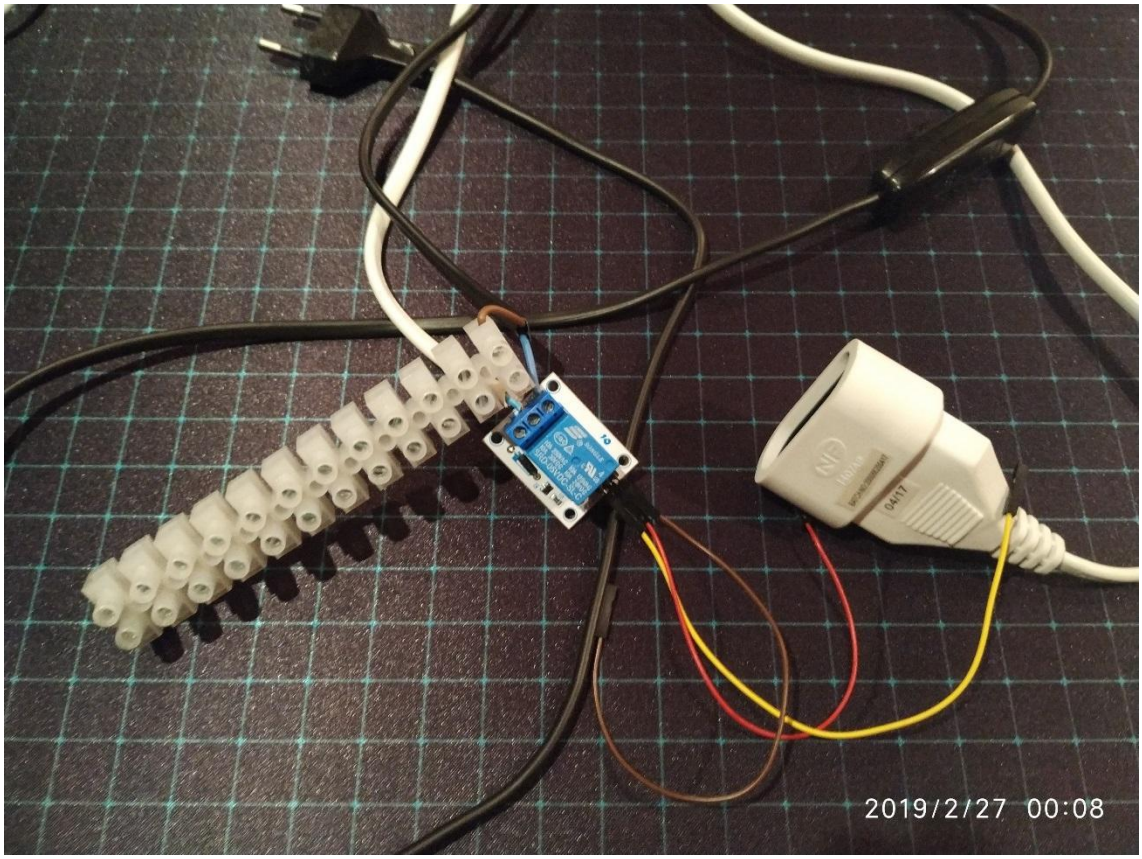


Figura 75 - Relé "ilumina-nos" quando temos resultados positivos

6.3 Conclusões

A vida hoje é mais rápida, as comunicações são instantâneas, e a interacção com pessoas do outro lado do mundo, assim como dos mercados globais, obrigam a tomar decisões a todo o momento, que têm impactos reais na nossa vida e na de que nos rodeiam.

Com a tecnologia a evoluir dia para dia e com a cada vez maior necessidade de informação e de decisões imediatas, há também a obrigação/necessidade de estarmos informados do mundo que nos rodeia, dos nossos concorrentes, dos nossos amigos, dos mercados, de tudo. É, portanto, cada vez mais necessária a OSINT, para melhorar a qualidade das nossas decisões em tempo útil.

Em poucos anos, a OSINT caminhou a passos de gigante e hoje todos reconhecem a sua importância, desde os serviços de informações/inteligência, aos órgãos de comunicação social, às forças de segurança e de investigação, aos militares e a todos aqueles que diariamente lutam na guerra dos negócios. Negócios estes, cada vez mais desleais e com competidores de todos os cantos do mundo.

Saber obter informações OSINT pode fazer a diferença entre ser burlado ou fazer um bom negócio, pode evitar a má opinião dos compradores, pode fazer reconhecer uma *fake new*, pode detectar concorrência desleal, detectar criminosos em fuga, detectar cartões e equipamentos roubados, detectar terroristas, reconhecer boas oportunidades, conhecer pessoas, evitar perigos.

Acreditamos que no futuro, o OSINT vai ser essencial em cada organização, e por experiência pessoal, temos visto isto. Para cada organização que tenho trabalhado, tenho oferecido serviços OSINT sob a forma de relatórios semanais, que possibilitaram por diversas vezes, a descoberta de ameaças à organização (manifestações, ataques informáticos com data, autor e hora definidos, opinião da sociedade quanto à empresa/Governo, entre outros).

Os desafios ao OSINT existem, embora a sua recolha seja feita em meios que por serem públicos e estarem ao dispôr de qualquer pessoa, podem ser utilizados sem receio de acusações penais ou outras. Os desafios decorrem da cada vez maior protecção dos sítios web em relação aos seus conteúdos, um pouco por causa da concorrência desleal, e como forma de forçar os visitantes a visitar o sítio web em vez de usar “atalhos”. Cada vez mais é difícil entrar num sítio web que não tenha publicidade, artifícios para nos introduzirem *malware*, cookies, identificadores de sessão, entre outros.

Acreditamos que o OSINT é um oceano ainda com muito para navegar e descobrir. O iKNOW não pretende ser revolucionário nem pretende dizer que o OSINT é superior a qualquer uma das formas de Intelligence existentes, simplesmente demonstrar que é possível fazer mais, desenvolver novas ferramentas, tirar partido do que a informação aberta nos dá, para “antever perigos e evitá-los”.

O iKNOW pode ser um pequeno passo para muitas ferramentas que ainda não existem (“trabalhos futuros”), e que poderão colmatar falhas nesta área. O trabalho aqui apresentado também pode auxiliar nas decisões quanto à escolha da arquitectura de rede e componentes.

Os testes e inquéritos efectuados com os utilizadores foram muito úteis e permitiram aferir com algum grau de confiança que a última versão responde ao que se quer e corrige algumas falhas, nomeadamente de usabilidade e curva de aprendizagem. Há, no entanto, ainda muito por fazer e melhorar. Algumas decisões podem não ter sido as melhores, mas houve um grande processo de criação, e uma grande aprendizagem ao longo destes 5 anos.

O autor acredita que como está, a ferramenta criada como prova de conceito, é escalável, resiliente, fácil de implementar e usar, boa como plataforma de estudo. Superou o inicialmente previsto e já é útil, auxiliando nos esforços do dia-a-dia para a obtenção e compilação de informações, ajudando também à elaboração de relatórios, permitindo obter notícias de sites que as ocultam, permitindo a localização de equipamentos furtados, entre outros.

APÊNDICES

1. OSINT – Evolução histórica

Como referido no capítulo de Ramos da inteligência, a origem dos diversos ramos de *intelligence* é militar e oriunda dos Estados Unidos, uma das razões pela qual a OSINT tem tido naquele país, tanta visibilidade, credibilidade e de onde se tem vindo a "espalhar".

A origem das fontes abertas remonta a vários séculos atrás (sem que haja uma data fixa ou aceite no geral), baseados na publicação e republicação de livros e publicações periódicas.

Segundo o artigo "*History of CIA Collection of Publicly Available Information (Open-Source Collection)"*[193], o OSINT teve origem durante a II Guerra Mundial, e era feito pelo departamento "*Foreign Broadcast Monitoring Service"*. Um dos motivos (e simultaneamente um dos receios) para a criação inicial do FBMS e ter ido para a alçada da CIA, eram o da propaganda estrangeira que invadia as fronteiras nacionais através da difusão via rádio. O mesmo site referencia que ainda hoje, existe este medo, usando o exemplo do recrutamento de futuros terroristas do ISIS/ISIL, recorrendo às redes sociais.

No entanto o termo OSINT só foi criado em 1980[194].

Em 1942, o departamento passa a designar-se por *Foreign Broadcast Intelligence Service*[195]*(FBIS), e mantém* esse nome e funções até 2005.

Mas antes disso, é importante referir que em meados de 1898, começa a haver uma difusão realmente em massa, da informação, através dos meios de massa tradicionais: radio, cinema, televisão, jornais, revistas. Esta época estendeu-se até cerca de 1994. A informação era criada e dirigida apenas numa via. Criador -> consumidor.

Esta difusão em massa já produzia muita informação e que chegava a muita gente.

Em Inglaterra, por esta altura, foi criado o *British Broadcasting Corporation (BBC, 1939 a 1949)*[196], uma agência governamental, de notícias OSINT que observava as informações(notícias) difundidas(propagadas) pelos órgãos de comunicação dos diferentes países, e "oferece" estas informações, depois de digeridas, como um serviço a si mesmo (Governo), a organizações comerciais (subscrições pagas) e ao cidadão comum (através dos seus programas noticiosos). Esta agência vai colaborar depois com as congéneres americanas

[193]Disponível online em https://www.lawfareblog.com/cias-new-guidelines-governing-publicly-available-information, acedido em 2018-01-03

[194] "The US military first coined the term OSINT in the late 1980s, arguing that a reform of intelligence was necessary to cope with the dynamic nature of informational requirements, especially at the tactical level on the battlefield", The Evolution of Open Source Intelligence (OSINT), Florian Schaurer and Jan Störge

[195]Disponível *online* em https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/foreign-broadcast-information-service/1-FBIS-Early-Beginnings.pdf

[196] Patrolling the Ether: US–UK Open Source Intelligence Cooperation and the BBC's Emergence as an Intelligence Agency, 1939–1948, Laura M. Calkins

partilhando as informações obtidas, até 1948. A divisão da BBC que faz a monitorização, tradução, etc, é a BBCM (*monitoring)*. A importância desta agência é muito grande, veja-se[197]:

- *Mediação entre presidentes Kennedy e Khrushchev relacionada com a crise de mísseis em Cuba;*
- *Informações iniciais relativas ao desastre nuclear da Ucrânia em 1986;*
- *Obtenção de informações locais e regionais onde o Reino Unido planeia ou intervém a níveis políticos e militares;*
- *…*

Esta mesma agência goza de excelente reputação. Ficam algumas características:

- Habilidade "sem rival" de correlacionar informação aparentemente irrelevante, conectando-a a diversas disciplinas e analistas de segurança;
- Ser a primeira a gerar alertas sobre teorias da conspiração, que podem ser de brincadeira ou de facto, serem mortais;
- Oferecer contexto e pensamentos críticos dos desenvolvimentos internacionais;
- Ser das poucas actualmente a combinar informações em suporte escrito e electrónico, e a permitir interacções sociais em sítios web;

Durante a *Guerra Fria* (1947 a 1991), a OSINT vai destarcar-se por ser a melhor fonte de informações, "*…Open sources eventually became "the leading source" of information about the adversaries', military capabilities and political intentions, including early warning and threat forecasting…*"

Em 11 de Setembro de 2001, com os ataques terroristas às torres gémeas, alguns serviços de informações são repensados para fazer face às novas ameaças e face à incapacidade de não ter prevenido tão importante ataque.

Em 2004/2005 os serviços do FBIS dão origem ao *National Intelligence's Open Source Center[198] (OSC),* que ficou responsável pela OSINT e pelas operações de filtrar, transcrever, traduzir, interpretar, processar e arquivar, todos os tipo de informações de media estrangeiros. Houve um relatório, entregue ao Presidente dos EUA(George W. Bush) e que deu origem a esta mudança)[199].

O relatório citado[200] anteriormente tinha vários pontos de muito interesse para todos aqueles que gostam do tema da *intelligence* mas aquele que nos interessa para este trabalho traz o seguinte texto, aqui colocado *ipsis verbis*:

"Create an Open Source Directorate within the CIA: We are convinced that analysts whouse o pen source information can be more effective than those who don't. Regrettably, however, the Intelligence Community does not have an entity that collects, processes, and makes available to analysts the mass of open source information that is available in the world today. We

---

[197] https://medium.com/@jonathanmarks/open-source-stupidity-the-threat-to-the-bbc-monitoring-service-deaaa9a393b4, 3-2-2017

[198]https://www.cia.gov/news-information/press-releases-statements/press-release-archive-2005/pr11082005.html

[199]Através das recomendações do relatório *Unclassified Version of the Report of the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction* , disponível online em https://www.gpo.gov/fdsys/pkg/GPO-WMD/pdf/GPO-WMD.pdf. Páginas 22 e 23.

[200]*Unclassified Version of the Report of the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction* , disponível online em https://www.gpo.gov/fdsys/pkg/GPO-WMD/pdf/GPO-WMD.pdf

*therefore recommend the creation of an Open Source Directorate at the CIA. The directorate's mission would be to deploy sophisticated information technology to make open source information available across the Community. This would, at a minimum, mean gathering and storing digital newspapers and periodicals that are available only temporarily on the Internet and giving Intelligence Community staff easy (and secure) access to Internet materials. In addition, because we believe that part of the problem is analyst resistance, not lack of collection, we recommend that some of the new analysts allocated to CIA be specially trained to use open sources and then to act as open source "evange-analysts" who can jumpstart the open source initiative by showing its value in addressing particular analytic problems. All of this, we believe, will help improve the Intelligence Community's surprisingly poor "feel" for cultural and political issues in the countries that concern policymakers most. The Open Source Directorate should also be the primary test bed for new information technology because the security constraints—while substantial—are lower for open source than for classified materia**l."***

Em 1 de Outubro de 2015, o Open Source Center(OSC) adopta o nome de *Open Source Enterprise(OSE)*[201]. Esta mudança do FBIS para o OSC/OSE não foi positiva para a comunidade que até então tinha algum retorno de informação. *"That decision was primarily due to the cost-prohibitive nature of updating the feed and in light of the broad accessibility of open source information on the Internet,".* No entanto, no sítio web da FAS[202], é referido que os serviços e relatórios gerados eram vistos com muita controvérsia pois consideravam-se inimigos da livre concorrência. *(Tendo este serviço terminado, a BBCM inglesa, pôde começar a vender este serviço, o que é interessante porque os fundos que lhes eram dados para este (BBCM) departamento foram cortados naquela altura... Uma feliz coincidência, evidente.)*

De 1994 a hoje, com a invenção e massificação da Internet e da informação, é possível a qualquer pessoa criar conteúdos e difundi-los instantaneamente para milhões de pessoas, sem necessidade de qualquer revisão/autorização de terceiros. Ou seja, uma relação Consumidor->consumidor, sem que, no entanto, deixem de existir os meios de comunicação de massa já existentes. Exemplo: jornais especializados, dissertações académicas, relatórios de empresas, diários pessoais, vídeos, fotos (incluindo metadados), mapas topográficos, manuais de utilização de equipamentos. Tudo online, ou de simples acesso.

Em Portugal, 2005, o Jornal de Negócios dizia "a *OSINT nos últimos cinco anos tem vindo a ser objecto de interesse crescente no ambiente dos serviços de informações civis e militares e das revistas académicas da área dos Intelligence Studies. A NATO, por exemplo, tem vindo a criar doutrina em torno do conceito desde os finais de 2001*"[203].

Também em Portugal, mas pela Defesa, Informações e segurança informática, é cada vez mais reconhecido o papel e importância do OSINT, vendo-se por exemplo, o crescente

---

[201] "*said CIA spokesperson Ryan Trapani. "OSE remains dedicated to collecting, analyzing, and disseminating publicly available information of intelligence value. The organization's new name reflects the broad relevance and scope of the open source mission*."", disponível online em https://fas.org/blogs/secrecy/2015/10/osc-ose/, 07-07-2016

[202] CIA Cuts Off Public Access to Its Translated News Reports, disponível online em https://fas.org/blogs/secrecy/2014/01/fbis-wnc/, 8-01-2014

[203] http://www.jornaldenegocios.pt/opiniao/detalhe/o_novo_conceito_de_osint

número de cursos [204] [205] [206] [207] [208], teses académicas [209] [210] [211], serviços, e consumidores a criar e a procurar informações deste tipo para conhecer perigos, evitá-los, descobrir oportunidades (*competitive intelligence* por exemplo), conhecer o meio ambiente, e prosperar. Do estudado e das entrevistas efectuadas com algumas fontes, aparentemente as nossas forças estão dotadas ou em vias, de ter capacidade de obtenção automatizada, sendo que fontes humanas já verificam OSINT diariamente como forma de conhecer perigos e evitá-los. Por exemplo, em SOC's, CSIRTS, serviços de informações e de investigação criminal.

Em termos salariais, na América as ofertas de trabalho pela CIA[212], têm à data de hoje (Janeiro de 2019), ordenados a rondar os $56,996 - $145,629 e mesmo assim, tal tem feito a mesma agência perder muitos agentes que saem para o sector privado. Em Portugal, analistas com as mesmas funções receberiam o ordenado de funcionário público, acrescido de 20% de disponibilidade e possível subsídio de risco. Ou seriam sub-contratados a empresas de outsourcing a preços ridiculamente altos. *Dá que pensar...*

O referido *Open Source Enterprise* está "de alguma forma" disponível na Internet, através de um site[213] (imagem do mesmo pode ser vista abaixo) mas apenas para alguns membros: Governo, Polícia, Serviços de Informações, e outros órgãos do Estado. Não deixa de ser caricato, a criação de *Open Source* para consumo interno, e com credenciação minima (e americana) de secreto. Se repararmos na imagem abaixo, podemos ver que a informação pode ser acedida por diversos grupos com acordos e também pelos empregados da já referida congénere inglesa BBCM. (*Curioso…)*

[204] *Advanced web exploitataion – human hacking* - https://www.cybers3c.pt/event/formacao-advanced-web-intelligence-human-exploitation/

[205] *Social Media Intelligence* - http://www.unl.pt/guia/2012/isegi/UNLGI_getUC?uc=42384

[206] Pós-Graduação em Inteligência Competitiva e Cibersegurança - https://www.ulusofona.pt/pos-graduacoes/inteligencia-competitiva-e-ciberseguranca

[207] Marketing Intelligence - https://www.flag.pt/masterclass/marketing-intelligence/

[208] Intelligence & Relações Internacionais no século XXI - https://www.ecs.uevora.pt/divulgacoes/cursos_livres/Intelligence-Relacoes-Internacionais-no-seculo-XXI

[209] Prova de Mestrado: *Threat Intelligence: Using Osint and Metrics to Enhance Siem Capabilities*

[210] Cyber Intelligence – obtenção de informações a partir de fontes abertas no ciberespaço - https://comum.rcaap.pt/bitstream/10400.26/8611/1/2013_12_09_OF_Disserta%C3%A7%C3%A3o_HFinal.pdf

[211] Prova de Mestrado "*Leveraging OSINT to Improve Threat Intelligence Quality*" - https://ciencias.ulisboa.pt/pt/evento/18-01-2019/prova-de-mestrado-leveraging-osint-to-improve-threat-intelligence-quality

[212]Disponível online em https://www.cia.gov/careers/opportunities/foreign-languages/open-source-exploitation-officer-1.html, acedido em 2018-01-03

[213]*Open Source Enterprise,* site da CIA para OSINT - disponível online em https://www.opensource.gov/public/content/*login*/*login*.fcc

Figura 76 - CIA: Open Source Intelligence mas só para "consumo interno"

2. *Fake news*: propagação, impacto e combate

Fake news é um termo estrangeiro muito utilizado hoje em dia, por toda a gente (incluindo presidente das maiores potências económicas e militares do mundo) para representar "notícias falsas". Informações que são espalhadas como sendo verdadeiras, especialmente pela Internet e redes sociais, mas que não são. Assentam fortemente em apelos dramáticos, emocionais e não tanto nos apelos factuais.

A perigosidade das fake news é muito grande. Uma pequena notícia dada por um site duvidoso na Internet, pode fazer um órgão de comunicação verídico, espalhar informação falsa par milhões de pessoas. Quem ouve fake news, pode ser levado a acreditar e comprar algo, pode amar ou detestar alguém, pode ser levado a denegrir ou a melhorar a sua imagem de alguém, ou de algo. Veja-se por exemplo as eleições americanas, tão polémicas por se suspeitar que tiveram mão russa nas eleições[214], podendo ter favorecido um candidato e denegrido outro. Ou seja, poderemos dizer que as fake news, poderão ser usadas como uma armam como propaganda, ilegalmente e sem ética, contra algo ou alguém, de uma forma sem controlo nem precedentes, especialmente se o alvo destas informações forem pessoas cujo desconhecimento em relação às matérias em causa, seja baixo.

O combate às fake news, é recente e de tal forma importante, que em 2016, diz o site da BBC[215], que os dicionários "*Oxford Dictionaries* ", declararam a palavra "Post-truth" (pós-verdade) como sendo a palavra do ano, não só devido à importância mediática das eleições, como de várias polémicas que se seguiram, entre elas, o papel das redes sociais, pelo seu crescente uso (2,000% de crescimento em relação à utilização da palavra relativamente a 2015). A importância da escolha de "post-truth" é enorme e pretende disseminar a ideia que é preciso combater as fake

[214] Eleições americanas, 2016, tendo ganho Donald Trump.
[215] https://www.bbc.com/news/uk-37995600

news, procurando a verdade. Também a Google encara as fake news como uma prioridade e criou um programa de educação para crianças[216] com este tema como objectivo.

Exemplo do perigo, obtido no sítio web da TVI24[217]:"*…para além dos EUA, também no Brasil as notícias falsas são recorrentes e têm consequências visíveis. Com 120 milhões de utilizadores do WhatsApp, os responsáveis pelas campanhas eleitorais viram aqui uma autêntica autoestrada livre de obstáculos para fazer passar a mensagem. É fácil, é barato e… pode dar milhões de votos."*

Ficam algumas formas, de conhecimento comum, de identificação e combate às notícias falsas. A informação que leu algures é fora do comum, é fantástica, está fora do padrão ou é contra aquilo que achava que era?

- Conteúdo total: de onde veio a informação tem a notícia completa ou foi apenas publicidade ou um chamariz? Podemos (e devemos) ler toda a informação.
- Fonte e referências: de onde veio a informação? É uma rede social? Tem hiperligações para um sítio web exterior que tem a informação no seu contexto? O sítio web é conhecido e bem-reputado? É um jornal online reputado, um jornal sensacionalista ou nem jornal é? Tem contactos ou forma de comprovar a informação? O sítio web é mesmo de notícias ou tem outras informações deste género, pode ser um sítio web de sátiras ou brincadeira?
- Autor: tem autor na notícia? O autor existe? É conhecido e aparece em outras notícias, são do mesmo género? Outros sítios web dizem a mesma coisa ou apenas este? Se o alvo da notícia é uma pessoa ou um grupo, poderá o autor ter algo contra, do tipo preconceito ou algo semelhante?
- Quando foi publicada esta informação? Há ali algum interesse para vender algo ou levar o leitor a fazer algo?
- Conhece alguém que possa comprovar ou desmentir o que viu?

Num artigo[218] do sítio web do jornal Observador, há aparentemente, diferenças entre as *fake news* de Portugal e de outros países. *«As 'fake news' políticas no resto da Europa lidam sobretudo com imigração, com refugiados, com a diferença, interculturalidade, o islamismo. Aqui em Portugal isso não faria qualquer sentido e, portanto, o tema muito mais presente neste tipo de 'sites' de desinformação é a corrupção, a forma como acusam políticos de terem roubado dinheiro», «Característica comum a estas páginas, em português, é, pois, o fraco peso da política nas publicações que fazem.»*.

A notícia do site acima, não refere, no entanto, que os sítios web que tentaram entrevistar por serem visivelmente sites de propagação de fake news, não são de origem portuguesa! Embora

[216] https://www.tecmundo.com.br/internet/143321-google-lanca-curso-portugues-criancas-identificarem-fake-news.htm

[217] https://tvi24.iol.pt/sociedade/notícias-falsas/estas-notícias-foram-fake-news-em-2018-e-muitos-acreditaram

[218] Como é o mundo clandestino dos "sites" em português associados às 'fake news', 23 de Fevereiro de 2019, https://observador.pt/2019/02/20/como-e-o-mundo-clandestino-dos-sites-em-portugues-associados-as-fake-news/

os títulos, as informações, etc, o sejam. Em novembro de 2018, a notícia[219] «*Fake news: sites portugueses com mais de dois milhões de seguidores»,* do Diário de Notícias, diz «*O negócio é rentável. A audiência dos sites de desinformação permite um retorno de milhares de euros pagos pela publicidade do Google. Só no Facebook, mais de dois milhões seguem estas páginas portuguesas...», «negócio das fake news é rentável. O que poucos portugueses conhecem é a real dimensão deste "mercado"»,* mas prossegue com alguns pontos interessantes:

- *…além dos crimes praticados pela desinformação propriamente dita (difamação, plágio, entre outros), esta atividade faz-se também à revelia das normas fiscais …*
- *…. Há mais de 40 páginas de desinformação portuguesas. É uma rede profissional. Ou, se quisermos, uma indústria de desinformação…*
- *…Segundo dados do sistema de análise usado pela redação do DN para avaliar a sua audiência em redes sociais, na última semana, dois dos três artigos escritos em Portugal com mais partilhas no Facebook foram feitos por páginas de desinformação…*
- *…três artigos dos cinco mais replicados vieram de* sites *que imitam órgãos de comunicação social, mas não estão registados na entidade reguladora (ERC), nem têm donos identificáveis…*
- *…A intenção de enganar é clara. Muitos destes* sites *apresentam-se como se fossem órgãos de comunicação social…*

As fake news não são todas iguais, evidentemente, e têm sido distinguidas alguns objectivos, podendo ser categorizadas como:

- Sátira – em 2013, o sítio web do JPN[220], tinha como título, "Redes sociais: sátira política ou guerra entre candidatos?", o que já é demonstrador da utilização das redes sociais para estes fins. Três anos antes da interferência nas redes sociais, aquando das eleições de 2016 para presidente.
- *Phishing* e chamarizes – uma notícia com um título fantástico acompanhados de um primeiro texto, que só quem clicar, consegue ver que afinal a noticia não é assim. Isto é muito comum, pouco ético e sem muito a fazer. Os objectivos são no geral, lucrar com as visitas e publicidade, com uma clara parecença com os emails de *phishing*. As imagens ou o texto não condizem com a verdadeira informação.
- *Conteúdo fabricado* – puras notícias falsas. Conteúdos políticos, racistas, propaganda, entre outros. Pode integrar-se também aqui as notícias cujo propósito é unicamente denegrir alguém, com mentiras e supotsos "factos" (falsos). *"**A estratégia destes 'sites' passa, muitas vezes, por construir uma "notícia" a partir de uma informação verdadeira**, mas, depois, é tirada uma conclusão excessiva ou fora do contexto. Ou pura e simplesmente trata-se de uma informação manipulada"* [221].
- Conteúdo manipulado – notícias verdadeiras são misturadas com notícias falsas, de modo a manipular a opinião de quem não está mesmo dentro do assunto

[219] Fake news: sites portugueses com mais de dois milhões de seguidores , 11 de Novembro de 2018, disponível em https://www.dn.pt/edicao-do-dia/11-nov-2018/interior/fake-news-sites-portugueses-com-mais-de-dois-milhoes-de-seguidores--10160885.html

[220] https://jpn.up.pt/2013/09/13/redes-sociais-satira-politica-ou-guerra-entre-candidatos/

[221] Como é o mundo clandestino dos "sites" em português associados às 'fake news', 23 de Fevereiro de 2019, https://observador.pt/2019/02/20/como-e-o-mundo-clandestino-dos-sites-em-portugues-associados-as-fake-news/

- Muitos outros… Desde conteúdos de pessoas, que não existem, de universidades fictícias, ou de universidades que nunca fizeram tais estudos, …

- Conteúdo fabricado

Segundo o site do DN, na notícia referida anteriormente, foram identificados e constam na imagem abaixo, os sites de desinformação mais seguidos no Facebook. Contêm mais de 2.5 milhões de seguidores, todas anónimas, a apresentar-se errada e falsamente como órgãos de comunicação social e nenhuma delas registada em Portugal (…).

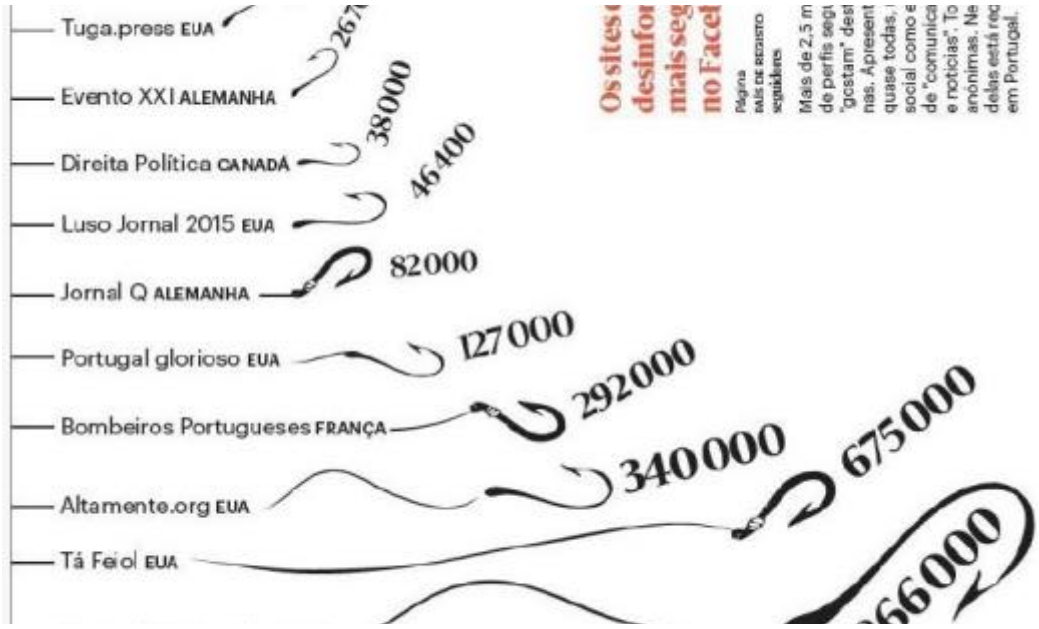


Figura 77 - Lista dos sites de desinformação mais seguidos no Facebook

3. Jornalistas - O contributo no combate às *fake news*

Os **jornalistas** (e órgãos de comunicação social) são os maiores produtores e consumidores de OSINT, seja porque produzem, seja porque, muitas das informações iniciais surgem desta via, ou de contactos que as souberam por fontes abertas.

Os jornalistas, têm nestes tempos de ouro da Internet, um manancial de informação quase ilimitado, com informações sobre tudo e todos. Saber distinguir o "trigo do joio", seleccionando e separando informação útil e verdadeira de falsa, nunca teve tanto impacto. A simplicidade com que se publica uma "notícia" falsa, juntamente com a dificuldade e complexidade de se encontrar o autor, aliado à falta de queixa em tempo útil, e à dificuldade de se compreender se a notícia é verdadeira ou não, visto que a difusão é grande e rápida, e o leitor pode não ter uma capacidade de análise muito grande (ou o tempo ou a vontade), fazem com que seja possível publicar o que quer que seja, sem que haja grande medo de consequências legais ou sociais. É aqui que o jornalismo tem de ter a responsabilidade de agir bem e éticamente a fim de proteger a sociedade em geral de ataques (*fake news* podem ser ataques de desinformação, contra-informação, propaganda, …) venham eles de onde vierem. A este

propósito, saiu um artigo[222] no jornal SOL, que dizia e contra-argumentava algumas ideias bastantes interessantes sobre os poderes das redes sociais, especialmente o Twitter enquanto plataforma de propaganda política e de disseminação de ideias. Vejamos, «*os tweets são propaganda e fazem doutrina*», ao qual se rebate que nem todos têm o poder de dizer em 250 caracteres algo que "mude o mundo", ou que tenham seguidores suficientes para o fazer, mas… existe a possibilidade e a ideia. Outra ideia que ali é rebatida e é interessante, é a seguinte (ainda relativamente ao Twitter) «*esta forma de comunicar é enganadora: serve para enganar jornalistas para que estes peguem nas mensagens e as ampliem*». Verdade?

Esperemos que os órgãos de comunicação social de hoje, estejam atentos a estas ameaças à verdade e cumpram o seu papel de informar, verificando e rectificando as notícias, com verdade e transparência. É certo que hoje qualquer um, pode mandar o que quiser para a Internet e ser lido por toda a gente, mas um jornalismo de qualidade não tem preço e tem (*ainda*) o seu lugar.

4. Defesa e forças militares - Geração e classificação de informações

"Um exército sem agentes secretos é um homem cego e surdo"
Sun Tzu, in "A Arte da Guerra"

A Defesa e forças militares são dos maiores interessados em informações/intelligence. Foi na defesa que nasceram muitos dos termos relativos a informações, usados hoje em dia não só nas Forças Armadas, mas também pelas forças de segurança e sociedade civil. A Defesa é obrigada a fazer o melhor uso possível das informações/intelligence de forma a ser o mais eficaz possível, já que um erro pode colocar vidas humanas em risco e a perda da soberania nacional. Serão de seguida introduzidas várias noções de informações/intelligence e a forma como a Defesa trata estes assuntos, muito semelhantes ao da sociedade civil, mas com uma importância maior.

Noção, Distinção e Classificação

Os militares consideraram que as informações não eram todas iguais e como tal, depois de distinguidas, classificaram-nas. Notícia por exemplo, é segundo o "Manual de Informações"[223], "qualquer facto, documento ou material susceptível de contribuir para um melhor conhecimento do inimigo actual ou potencial, ou da área de operações", com uma classificação que varia em função do seu grau de confiança (confiança na fonte, possibilidade de ser verdadeira comparando o que se conhece da noticia com o conhecimento da própria pessoa). Pode depois ser processada e arquivada para uso posterior, ou não.

[222] "A doutrina socialista sobre as *fake news* – uma polémica quixotesca", Sofia Vala Rocha, 2 de Março de 2019, disponível online em https://scontent.flis9-1.fna.fbcdn.net/v/t1.0-9/53295594_10215904682423763_5837736694232645632_n.jpg?_nc_cat=107&_nc_ht=scontent.flis9-1.fna&oh=84a245011670b9587e6da7dd3e812a92&oe=5D1FA414

[223] Manual de Informações, EME, pg 6

O termo "informação", é utilizado pelos militares como sendo o resultado/conhecimento obtido em função de um estudo (pesquisa, estudo e interpretação) de todas as notícias relacionadas com o assunto.

Já para "informações", a noção envolve o conhecimento de informação, mas de uma forma extrapolativa, que envolve saber porque aconteceu, por quem, se pode acontecer de novo e quando. Estas "informações" também conhecidas por intelligence e são usadas neste trabalho como forma de distinção de informações no âmbito não-militar.

Classificação das informações quanto à utilização

As informações criadas pela Defesa, podem classificar-se em estratégicas, operacionais ou tácticas, consoante a sua utilização posterior. São definidas como:

- Estratégicas, aquelas que dão origem a decisões estratégicas políticas e militares. Alvo: Governo e Ministérios
- Operacionais, aquelas que são recolhidas e necessárias para o planeamento de operações. Alvo: Comandos militares.
- Tácticas, aquelas que são recolhidas e projectadas para o desenrolar imediato de operações. Tem como destino os lideres/comandantes

Qualquer uma das informações acima, tem em conta a máxima premissa de que só deve ter acesso à informação quem realmente possa precisar dela para cumprir a sua missão.

Consoante a ameaça (*qualquer acontecimento ou acção em curso ou previsível que contraria a consecução de um objectivo e que normalmente é causadora de danos materiais ou morais*[224]), a Defesa, que é composta por diversas forças militares deve tomar a melhor resposta para ser o mais eficaz possível. Por exemplo, uma missão que envolva mar vai com certeza incluir a Marinha. Uma missão que inclua a exfiltração de tropas ameaças ou cercadas vai envolver a força aérea, possivelmente Comandos ou Rangers de Operações Especiais, quiçá também Marinha. Diferentes problemas requerem diferentes abordagens e estratégias, daí também a opção das informações OSINT ser tão importantes.

Uma missão de paz, que traga tropas de um país amigo vai com certeza ser melhor recebida do que tropas de um país remoto ou com quem tenha havido disputas no passado.

A Defesa/Estado-Maior General das Forças Armadas, tem também no seu interior, diversos organismos que procedem à colecta e disseminação de informações, ainda que apenas interiormente e no seu âmbito, como por exemplo, o Centro de Informações e Segurança Militares (CISMIL). À unidade que trata das informações já foram dados diversos nomes tais como, 2ª Divisão do EMGFA/DINFO/SIED/SIM/DIMIL. Actualmente a produção de informações militares está a cargo do SIEDM (produção de informações ao nível estratégico) e da DIMIL.

A constituição do exército português (obtido no próprio site do exército):

[224] Noção de ameaça: Gen Abel Cabral Couto, Elementos de Estratégia, I Vol, Edição do IAEM, 1988, pg 328.

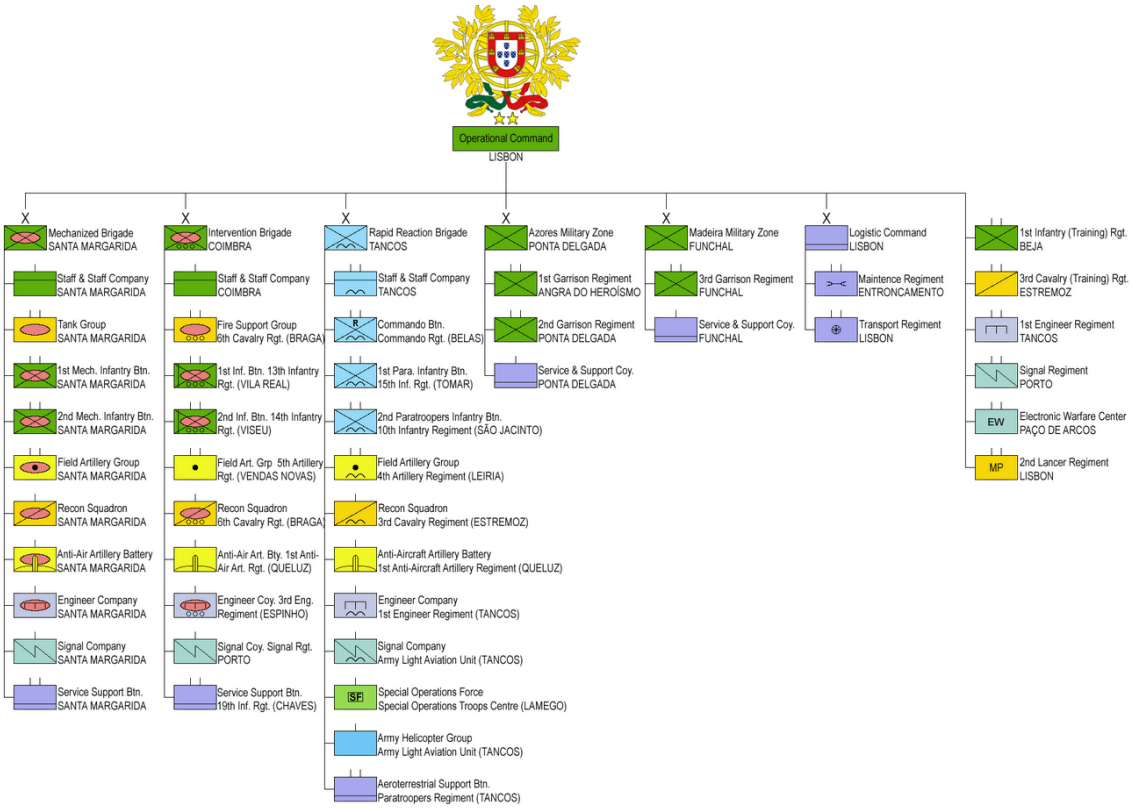


Figura 78 - Mapa de localização dos quartéis e tropas portuguesas

Nacionalmente, a Defesa organiza acções de sensibilização para a temática ciber, oferecendo alguns cursos pelo Instituto de Defesa Nacional. Nestes, foi possível aprender que em casos extremos de ameaça ou ataque cibernético, as Forças Armadas Portuguesas (assim como outras no mundo), podem ser activadas e chamadas a actuar fisicamente.

Lá fora, também os incidentes no ambiente ciber podem escalar e tomar proporções reais. Ficam dois exemplos recentes:

- Em 2019, Israel bombardeou um edifício que se julgava ser a base de operações cibernéticas de um Estado rival, que estava a atacar informaticamente Israel.
- Nos Estados Unidos,o Pentágono tem uma estratégia[225] de ciberguerra, que admite ações "preventivas" . Esta estratégia tem em conta as várias gravidades que podem levar as Forças armadas adquele país a retaliar num ciberataque relativamente a uma hierarquia de gravidade. Se houver um ataque que atinja 2% dos sistemas norte-americanos, poderá haver uma resposta concertada, liderada pelo Pentágono e pelo seu Ciber Comando militar.

O OSINT pode ser utilizado por serviços de informações para tentar averiguar se uma informação é ou não credível, sem colocar meios humanos no terreno, podendo preservar assim os seus recursos, quer humanos quer financeiros.

[225] Estratégia Norte-americana de ciberguerra, disponível online em http://tek.sapo.pt/notícias/internet/o_pentagono_tem_uma_nova_estrategia_de_ciberg_1439114.html

Um dos casos mais conhecidos de OSINT é mostrado diariamente na televisão, através do programa "*catfish*" que tenta ajudar pessoas, arranjando encontros e desmascarando perfis *online*. Não desvendam muita informação do que fazem para chegar aos resultados, mas é possível fazer o mesmo através das técnicas abaixo.

5. OSINT – Casos práticos de investigação

5.1 Descobrir se a fotografia de um perfil é real

Algumas redes sociais permitem carregar imagens, retirando os metadados[226], o que por si só já é bom pois dá alguma privacidade sem o utilizador se preocupar com isso. No entanto é possível utilizar essas mesmas imagens para procurar na internet em alguns *sites*, se essas imagens são ou foram vistas em outros sites. Um dos melhores *sites* da área é o TinEye[227]. Obtendo uma imagem do Google, de uma importante personalidade portuguesa e colocando (copiando) a *url* da mesma no *site* referido foi possível obter uma listagem de sites onde é utilizada:

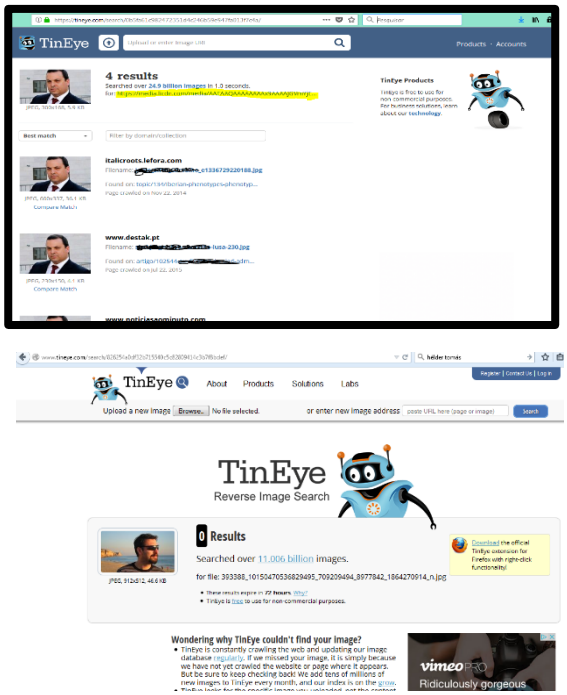


Figura 79 - OSINT - Caso prático- verificar foto de perfil

Nota: o nome da pessoa e sua função foram apagados propositadamente de forma a não violar a privacidade, ainda que estejamos a recorrer a fontes abertas.

O resultado obtido também era possível com fotos do *Facebook*, *LinkedIn*, entre outros. Este site permite saber quantas vezes aparece x foto. Se aparecer em muitas é provável que o nosso "amigo" da nossa rede social possa não ser quem aparenta ser.

5.2 Embarcações civis

A possibilidade de se obter informações sobre embarcações, sejam elas de pesca, recreio, militares, é sem dúvida algo que embora possa não ter uma importância prática para quem está a estudar OSINT, é de uma importância muito grande para quem faz a segurança física e a monitorização de cabos submarinos.

---

[226]Metadados - dados que não são a informação em si, mas um conjunto de informação que nos pode indicar a marca da máquina fotográfica, coordenadas gps, qual a impressora onde foi imprimido x documento, hora, data, quem foi, entre outros

[227]TineEye disponível online em https://tineye.com

5.3 Embarcações militares

Entrando no *site* do MarineTraffic[228], com um pequeno zoom sobre Lisboa, é possível observar que existem bastantes embarcações, dadas pelas cores e formas. Fazendo mais um *zoom* e clicando em cima de NRP Vasco da Gama, podemos confirmar se é o barco pretendido. A quantidade de informações é bastante grande e útil.

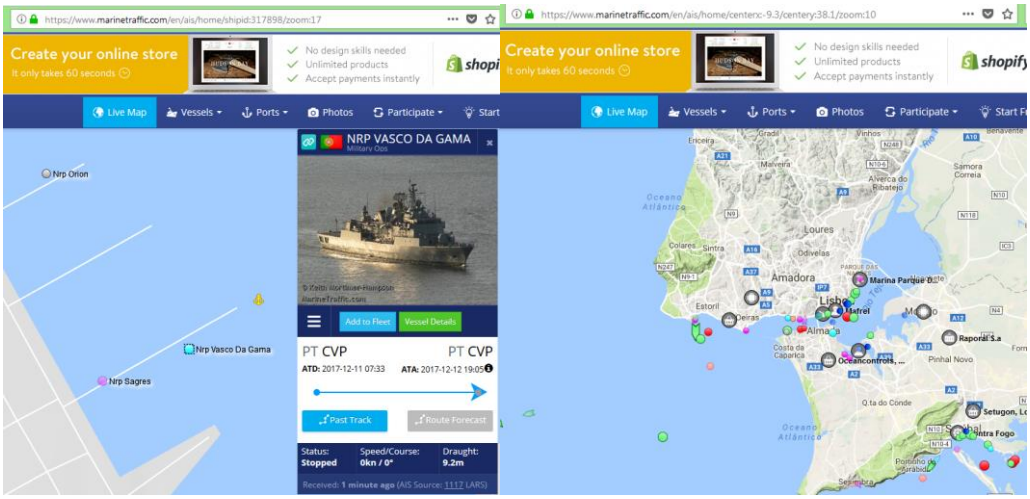


Figura 80 - OSINT - Procurar barcos de guerra no rio Tejo

5.4 Cabos submarinos

Recentemente, o autor fez uma visita a um centro de amarração de cabos submarinos[229]. Foi referido[230], que quando existem problemas de falta de conectividade, cabos danificados ou mesmo cortados, que verificam se naquela zona houve barcos. Pode não ser de forma intencional, mas as âncoras e as redes, podem levantar e/ou danificar os cabos de dados.

É possível verificar via OSINT, quem estava naquela zona, servindo também para responsabilizar, pedir indemnizações ou pelo menos, saber se o motivo poderá ter sido causa humana, e sendo, quem foi, e se propositado ou não.

[228]Disponível online em https://www.marinetraffic.com/
[229] Visita ao Centro de amarração de cabos submarinos, Sesimbra, 2019-06-06
[230] Embora sem querer atribuir palavras exactas ou colocar o nome do senhor

5.5 Aeronaves a sobrevoar-nos

Pretende-se saber o nome dos aviões que estão a sobrevoar Lisboa neste momento. Entrando no site Flight Radar 24[231], é possível verificar em tempo real, o nome do avião, a sua fotografia e o trajecto. Clicando em cima foi possível observar o tipo de avião, a hora prevista de chegada, de ontem vem e a foto.

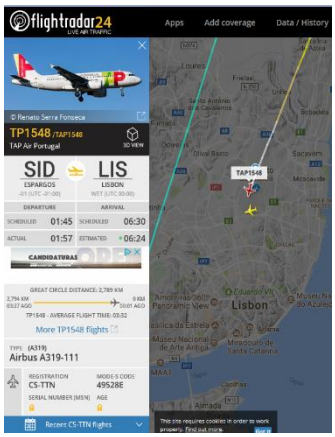
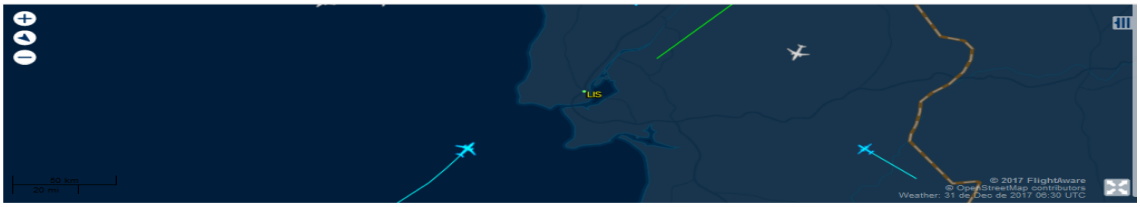


Figura 81 - Procurar aviões em tempo real

5.6 Em tempo real, nomes e horários de aviões

O site acima do Flight Radar é muito bom, mas para a ferramenta criada e para obtenção de informação, o Flight Aware[232] é melhor. Mapas aéreos ligados com o Google Maps, assim como os horários num formato bastante simples, para serem percorridos automaticamente por *crawlers/scrapers/spiders* (ou o nosso IKNOW).



Comprehensive flight data for LPPT is available to qualified aviation industry professionals. Learn more about FBO ToolBox

CHEGADAS (MAIS)

Número de voo	Tipo	Origem	Descolagem	Chegada
TAP1548	A319	Int'l Amílcar Cabral (SID)	01:57 -01	06:24 WET
DTA650	B773	Quatro de Fevereiro (LAD)	00:14 WAT	06:11 WET
TAP58	A332	Int'l de Brasília (BSB)	19:30 -02	06:11 WET
TAP104	A332	Int'l de Belo Horizonte-Confins (CNF)	19:10 -02	05:56 WET
TAP286	A343	Quatro de Fevereiro (LAD)	23:20 WAT	05:50 WET
TAP228	A332	Int'l de Miami (MIA)	17:40 EST	05:34 WET
TAP70	A332	Int'l do Rio de Janeiro-Galeão (GIG)	18:22 -02	05:20 WET
TAP202	A332	Int'l de Newark (EWR)	18:21 EST	05:04 WET
TAP1478	A320	Osvaldo Vieira Int'l (OXB)	00:53 GMT	04:44 WET
TAP1528	A320	Int'l Kotoka (ACC)	23:10 GMT	04:08 WET
MMZ2614	B763	Int'l Hato (CUR)	21:20 AST	03:03 WET

PARTIDAS

Número de voo	Tipo	Destino
EZY7687	A319	Basle-Mulhouse (EAP)
KLM1692	B738	Amsterdão Schiphol (AMS)
TAP1310	A319	Budapest Ferenc Liszt International Air
TAP13	A332	Int'l do Recife (REC)
TAP75	A332	Int'l do Rio de Janeiro-Galeão (GIG)
LZB564	A319	Sofia Airport (SOF)
TAP87	A343	Int'l de São Paulo-Guarulhos (GRU)
TAP1507	A320	Port Bouet (Felix Houphouet Boigny Int'l
TAP796	A320	Helsinquia-Vantaa (HEL)
TAP1324	A319	Int'l Henri Coandă (OTP)
TAP1230	A320	Int'l Domodedovo (DME)

Figura 82 - OSINT - Aviões - horários e informações de voo

5.7 Investigar sites e ficheiros por vestígios de malware

[231]Disponível online em https://www.flightradar24.com
[232]Disponível online em https://flightaware.com/live

Existem hoje ferramentas que são muito úteis para investigar por diferentes tipos de *malware* e que nos poupam o trabalho de dissecar ficheiros executáveis, assim como poupar muito tempo. Um site muito útil é o Virus Total[233]. Podemos enviar um ficheiro e o site devolve informação muito útil, tal como:

- A avaliação do ficheiro ou sítio web, por 70 motores de antivírus distintos
- O endereço IP do *site* que está a ser verificado
- Possíveis resultados anteriores
- A votação por parte de outros utilizadores que já testaram o site ou o ficheiro suspeito

O *site urlquery*[234] é também um site muito útil na hora de investigar pela credibilidade e conteúdo de um determinado site. Sem abrirmos o sítio web suspeito, podemos abri-lo utilizando uma ferramenta *online* que analisa o site e seu conteúdo, mostrando-nos uma imagem do que teríamos utilizando o nosso navegador de Internet. Mostra-nos também alarmes que o ficheiro levantaria num sistema com "Detecção de Intrusões". Este site em conjunto com o virustotal.com antes referido, permite analisar um site duvidoso sem colocarmos em risco o nosso sistema nem pôr em causa a nossa privacidade (Ip's).
Podemos verificar:

- o número de ligações com outros sites
- a credibilidade (se consta ou não das *blacklists*)
- o número de scripts executados e seu conteúdo
- o IP do servidor do site, a existência de alertas
- outros endereços de sítios web alojados no mesmo endereço
- dados DNS e do alojamento
- comunicações desse sítio web para outros sítios web, através de um gráfico com o caminho e saltos desde o sítio web inserido até à informação finalmente mostrada

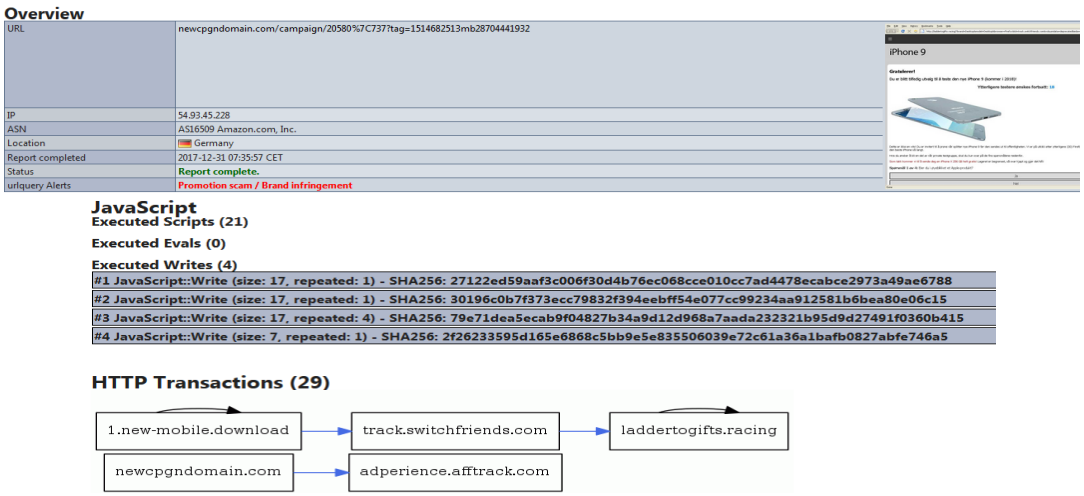


Figura 83 - OSINT - Verificação de site potencialmente maliciosos com UrlQuery

[233]Disponível online em https://www.virustotal.com
[234]Disponível online em http://urlquery.net/

6. iknow 1.0 - Apresentação e tutorial da ferramenta

6.1 Página de entrada

O utilizador, ao inserir a URL do *site,* depara-se com um *login*/formulário Web onde pode inserir nome e palavra-passe. É também mostrada alguma informação sobre utilizador, não só para efeitos de testes da ferramenta, mas também para de algum modo, avisar que o utilizador pode estar a mostrar dados que não quer (segurança do utilizador aqui em destaque). (“O que sei sobre si” é uma das formas que os donos do site podem fazer OSINT e saber/distinguir quem visita o seu site.) Na prática, foi utilizada esta informação para em testes, aceder ao servidor e sabermos se o nosso endereço IP público era o real, o da VPN ou o da rede TOR.

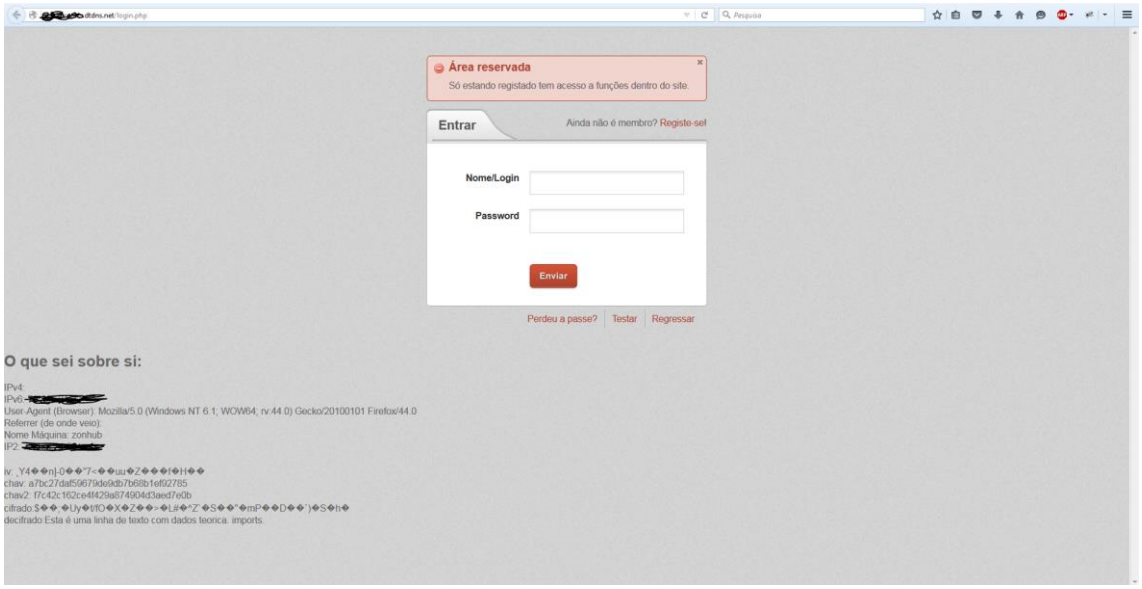


Figura 84 - iKnow - Página de entrada

Nesta página pode ser feita a autenticação ou registada uma nova conta.

Figura 85 - Página de entrada - Registo de utilizador

Figura 86 - Página de entrada - Autenticação

6.2 Scraping

Uma das secções mais importantes do *site*, é possível observar aqui sem mais demoras, o potencial do que é, e para que serve o iKnow. Não se pretende baixar todo o *site* e/ou suas informações, mas seleccionar apenas o que queremos para filtrar cada vez mais e obtermos informação. O OSINT como antes referido não serve para muito, se, tendo todas as informações, não formos capazes de as utilizar quando for necessário.

Figura 87 - Menu da secção de scraping

6.2.1 Europol – Obtenção da lista de pessoas procuradas

O site da Europol[235] contém informações sobre a própria instituição, ofertas de trabalho, objectivos, alertas vários que afectam os cidadãos europeus entre outros. Contém ainda uma lista das pessoas procuradas. Esta lista contém nome, foto, algumas informações sobre a pessoa e o crime cometido. A lista, no entanto, não nos permite fazer download da foto, e se o tentarmos obtemos uma imagem vazia. O iKNOW permite fazer *scrape* ao site de forma a baixarmos todas as informações de cada pessoa procurada, e ainda a dita foto protegida (obscurecida apenas no código).

Nota: se a foto estivesse protegida, o facto de quebrarmos a protecção invalidaria a informação como OSINT. Não é o caso.

[235] Site da Europol em https://www.europol.europa.eu/

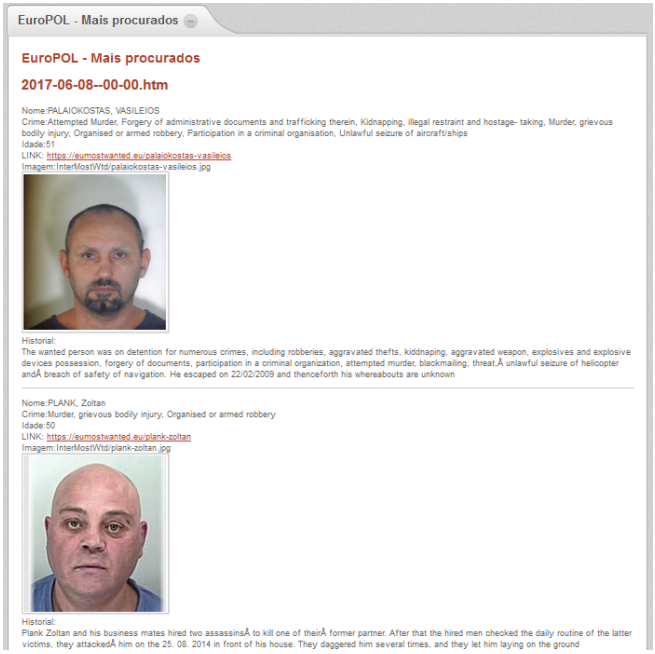


Figura 88 - Europol -Lista de pessoas procuradas. Foto, nome, crime, idade, …

6.2.2 Capas de jornais – Obtenção das capas

Nota: durante o *scrap* a diversos *sites* secundários de notícias, foram bastantes as vezes que os sites desapareceram ou simplesmente o seu conteúdo. Eis um deles, e do qual inicialmente obtia as capas dos jornais:

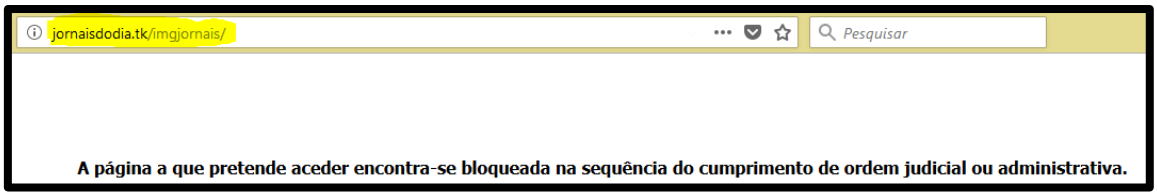


Figura 89 - Dificuldades: Site de capas de jornais foi bloqueado

Estudando como *site* é feito, é possível, no entanto, ir buscar a imagem ou conteúdo que se pretende. Imagem de capa de jornal abaixo[236] é um exemplo disso.

Figura 90 - Scrap a capas de jornais de sites de notícias

6.2.3 Sitio web de jornal – Obtenção de notícias, imagens e datas

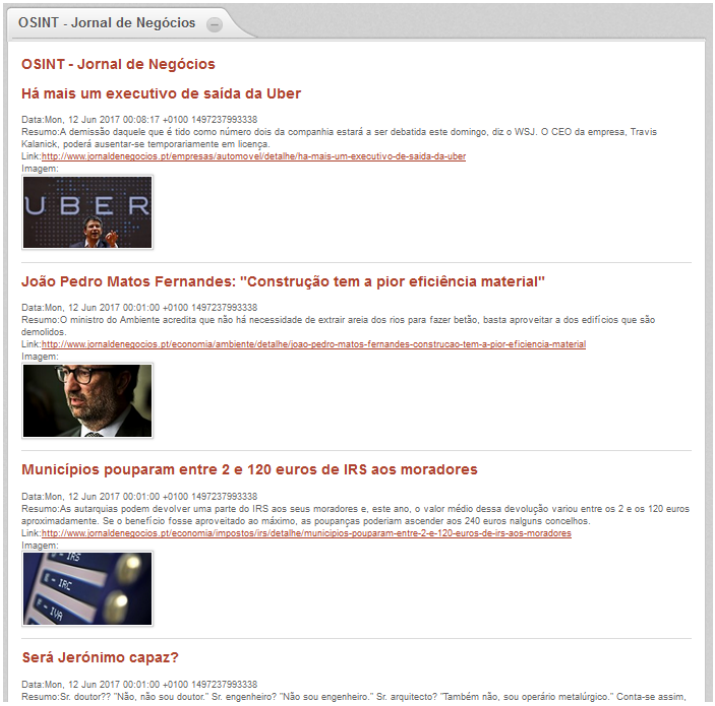


Figura 91 - Scrap ao Jornal de negócios – obtenção de data, resumo, links e imagem

[236]Capa do Jornal de Notícias. 31-12-2017

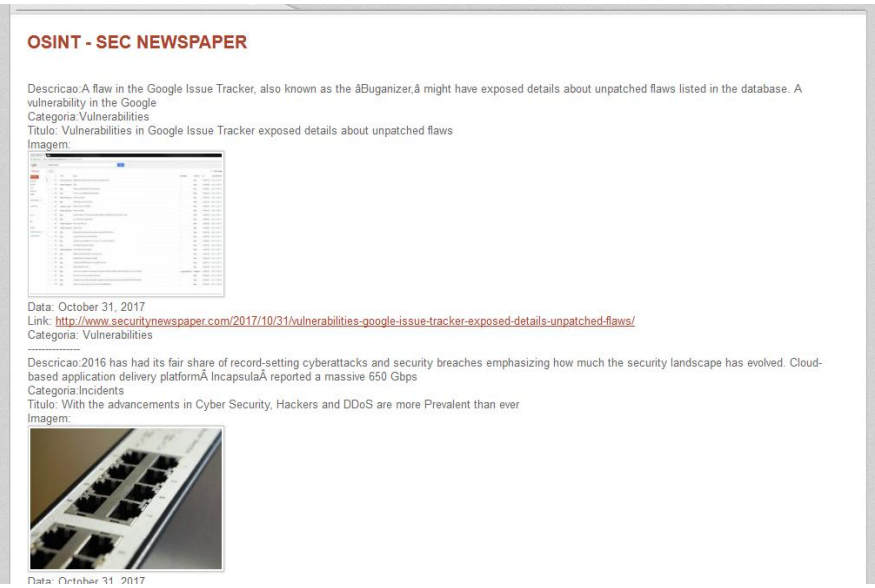


Figura 92 - Scrap ao site de notícias e ataques informáticos - security newspaper

6.3 Operações

Esta secção é/era o cerne da ferramenta. Nela, introduzimos todos os dados que pretendemos procurar. Nomeadamente, o URL do *site*, palavras-chave e/ou cartões de crédito. Define-se também aqui a data que a pesquisa deve ser feita assim como a sua frequência e tempo.

Estas pesquisas são pessoais e diferentes utilizadores só vêm os seus dados e estatísticas gerais.

As operações estão disponíveis primariamente em três opções, embora existam outras, apenas acessíveis no momento de edição. Cada categoria tem um menu diferente.

Figura 93 - Operações - Menu

Listagem de operações do utilizador em uso. Pode ser editada qualquer operação desde que a mesma ainda não tenha sido executada. Caso tenha sido, é possível visualizar ou apagar.

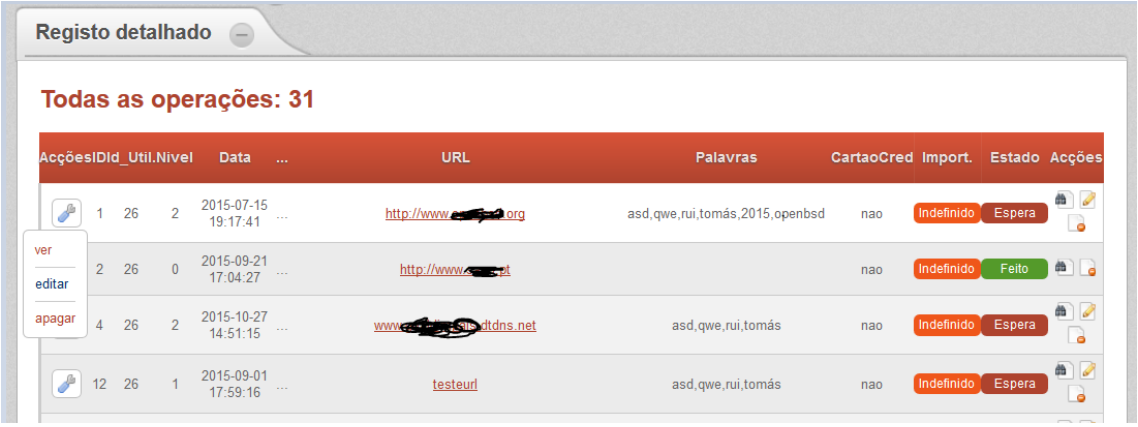


Figura 94 - Operações - Listagem total com menus dinâmicos

Cada utilizador poderá ver no final da listagem das suas operações, estatísticas sobre as operações totais do site. Esta característica é apenas informativa. Ninguém pode ver os dados de outros utilizadores.

Figura 95 - Operações - Listagem Parcial, Estatísticas pessoais e gerais

Exemplo de introdução de operações. Uma operação é um objectivo. Podemos querer saber se o nosso nome apareceu em x site. Para tal preenchemos o formulário e indicamos a data de início e a frequência em minutos para ser executada. Imaginando que o nosso nome aparece, a aplicação deve gravar a página total onde foi detectado o nome e enviá-la para o utilizador. CASO seja detectado algo, tudo é baixado e as imagens são analisadas por potenciais dados GPS que também serão acrescentados ao relatório enviado.

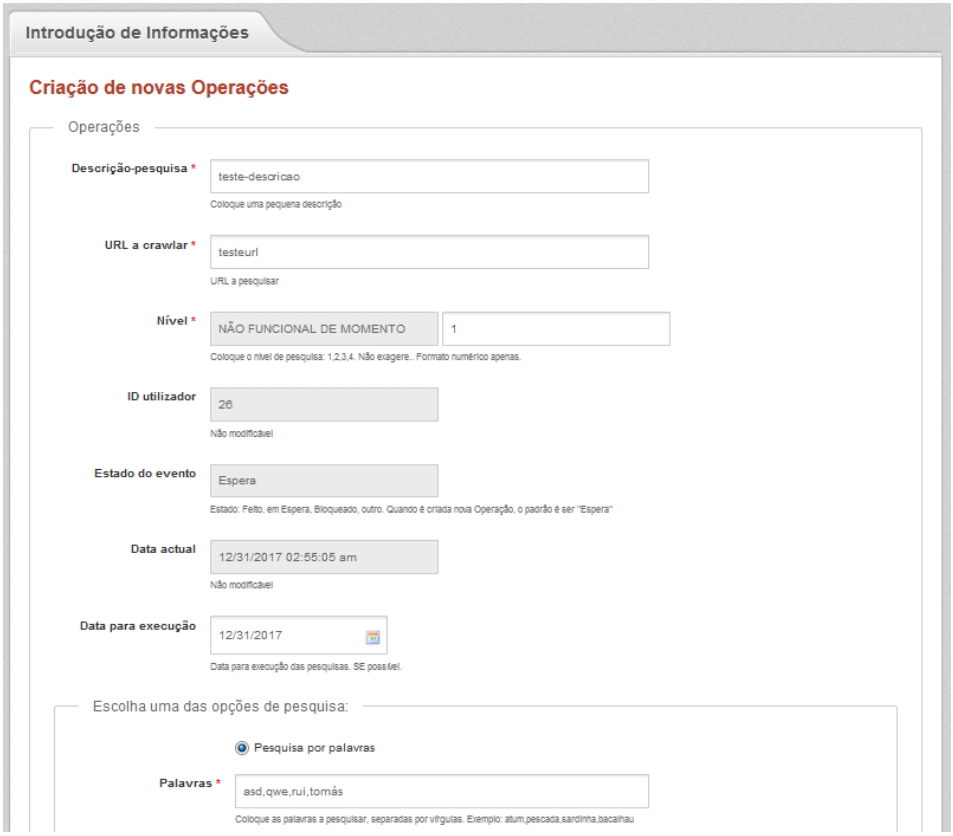


Figura 96 - Operações – Introdução

A imagem seguinte mostra uma página criada especificamente para ser impressa com todas as operações. A listagem mostra os dados básicos, a importância, entre outros.

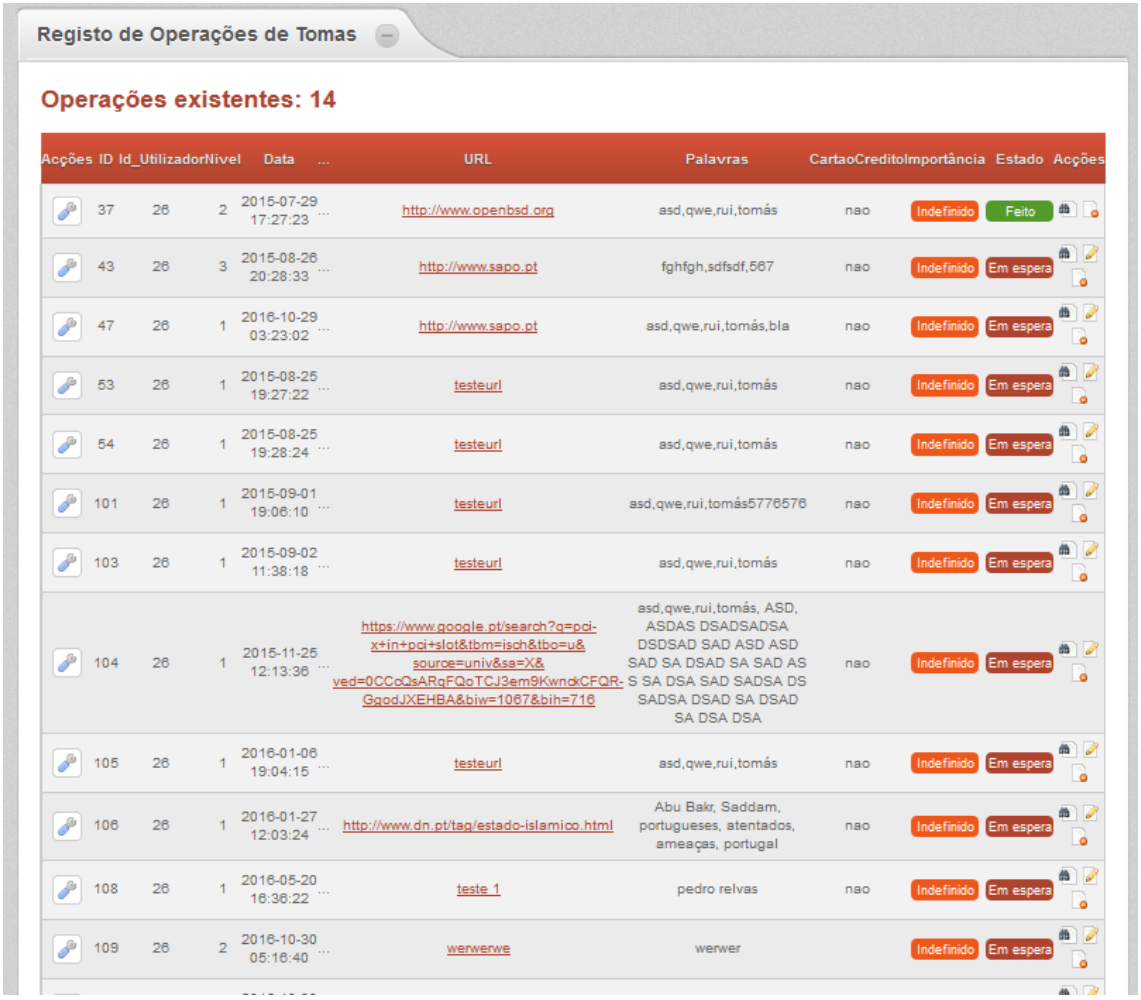


Figura 97 - Página de Listagem para impressão

Todos os resultados das buscas são aqui mostrados. A imagem mostra quatro resultados, decorrentes de duas buscas/operações (OP 37 e OP 3). O nome do ficheiro é o mesmo já que as páginas eram as mesmas. A busca (OP) 37 foi corrida três vezes e a busca 3, uma vez.

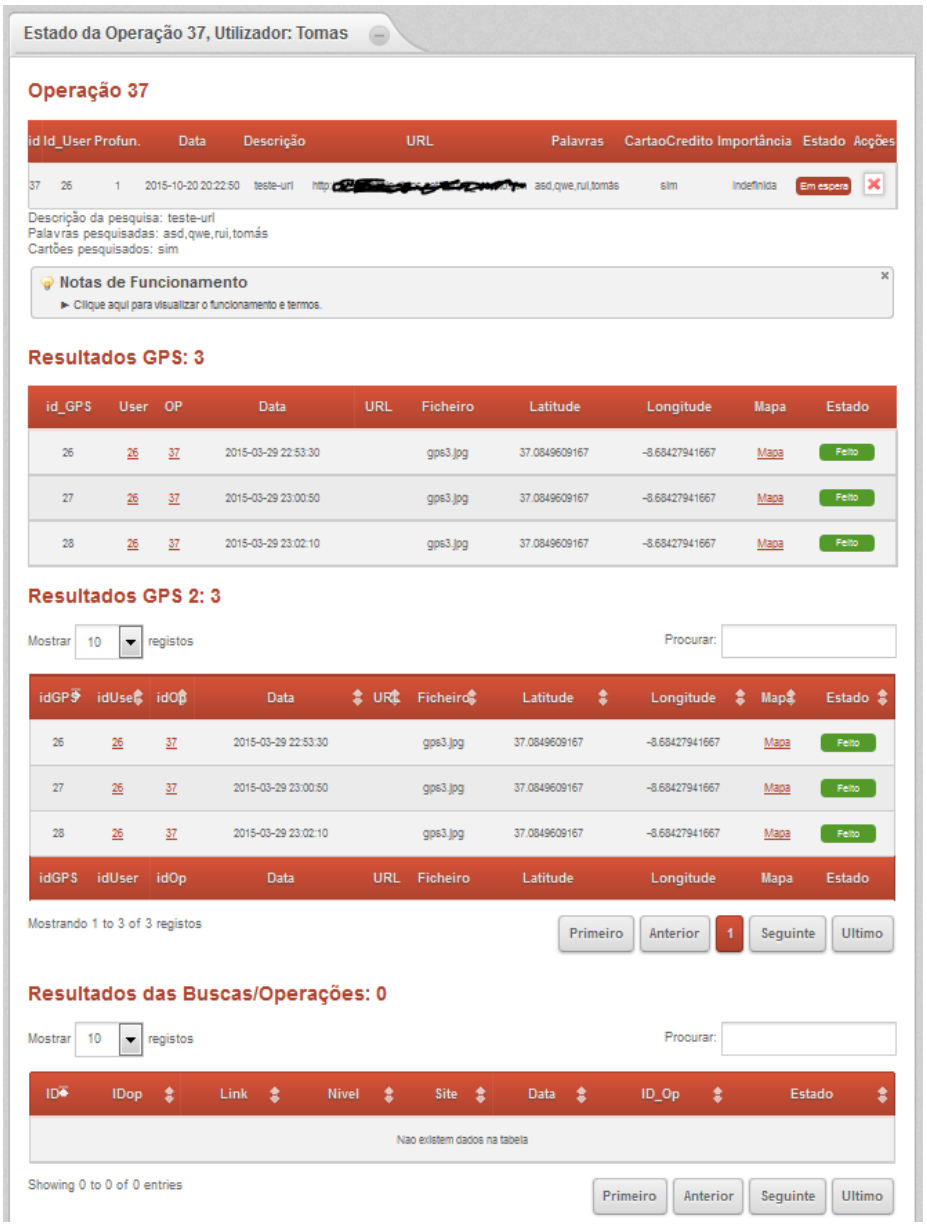


Figura 98 - Resultados - Listagem de resultados

6.3.1 Metadados de imagens

Clicando em cima da hiperligação "Mapa", a aplicação abre-nos o navegador de internet no *site* do *GoogleMaps*, mostrando-nos o local.

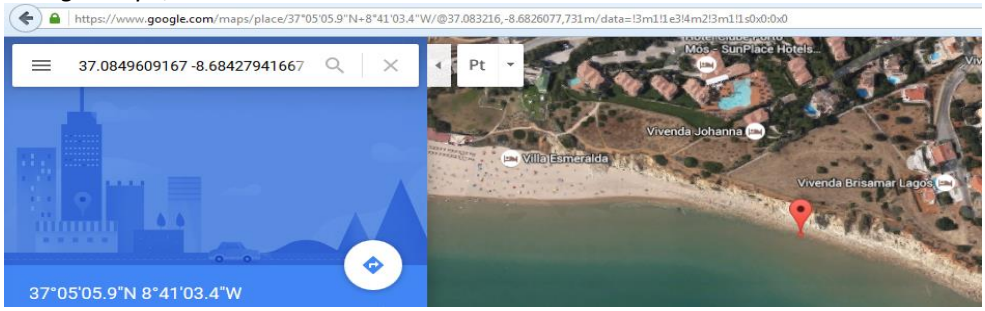


Figura 99 - Resultados - Coordenadas obtidas de imagem no Google Maps

6.4 Utilizadores

Cada utilizador é listado nesta secção e podem ser vistos alguns pormenores sobre o mesmo.

Clicando em utilizadores podemos ver uma listagem, diferenciando utilizadores e um gráfico com actividades de utilização.

Figura 100 - Utilizadores - Listagem, Estatísticas e Bloqueios

A área pessoal pode ser actualizada com a chave pública para envio dos resultados. Caso esta chave não exista, o *mail* é enviado sem qualquer cifra. Esta secção, futuramente, apenas vai mostrar pormenores básicos dos utilizadores. Apenas um administrador poderá ver tudo. A página mostra também alguns dados sobre o utilizador, à semelhança da página de autenticação e uma caixa mostrando se o utilizador é "válido" ou não (caso tenha sido bloqueado).

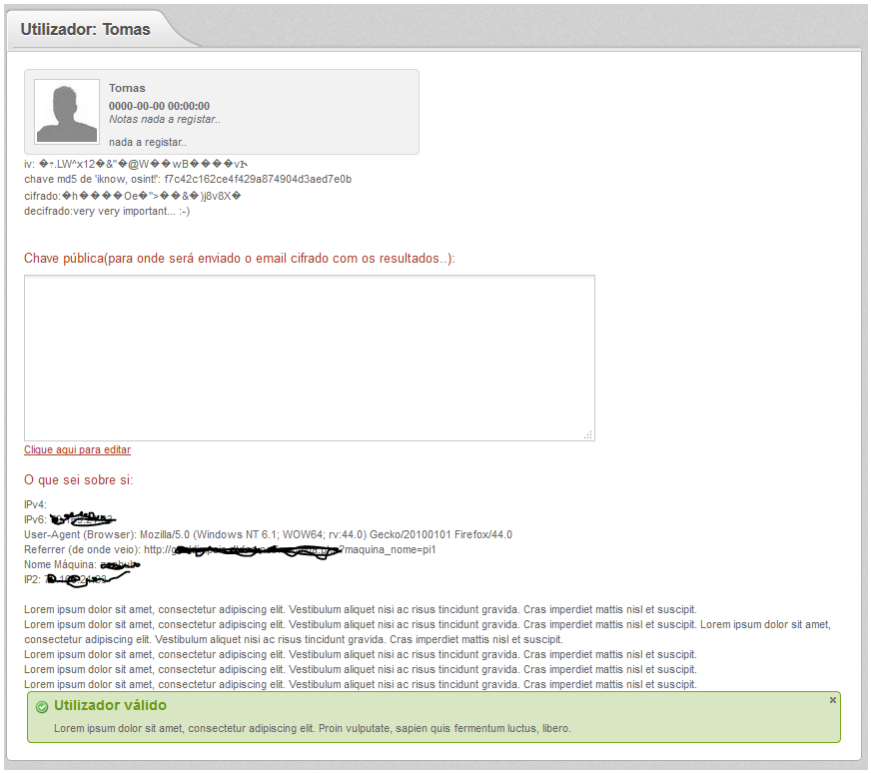


Figura 101 - Utilizadores - Ver utilizador

6.5 Gráficos, Estatísticas e Visitas

Os gráficos mostrados são feitos em *javascript* e actualmente não mostram nada real. Pretende-se futuramente encher com dados úteis e/ou métricas de utilização.

As estatísticas mostradas são reais, tais como os visitantes do site, operações de cada utilizador, etc. Calculam no carregamento da página, diversos itens que podem ser observados abaixo. No geral e no utilizador actual.

Figura 102 - Gráficos e Estatísticas - Estatísticas

Informações sobre o utilizador actual: endereço *ip*, *user-agent* utilizado pelo *browser*, de onde veio, nome da máquina e país. A listagem abaixo mostra o total de visitantes, a data, página de onde vieram, e *user-agent* utilizado.

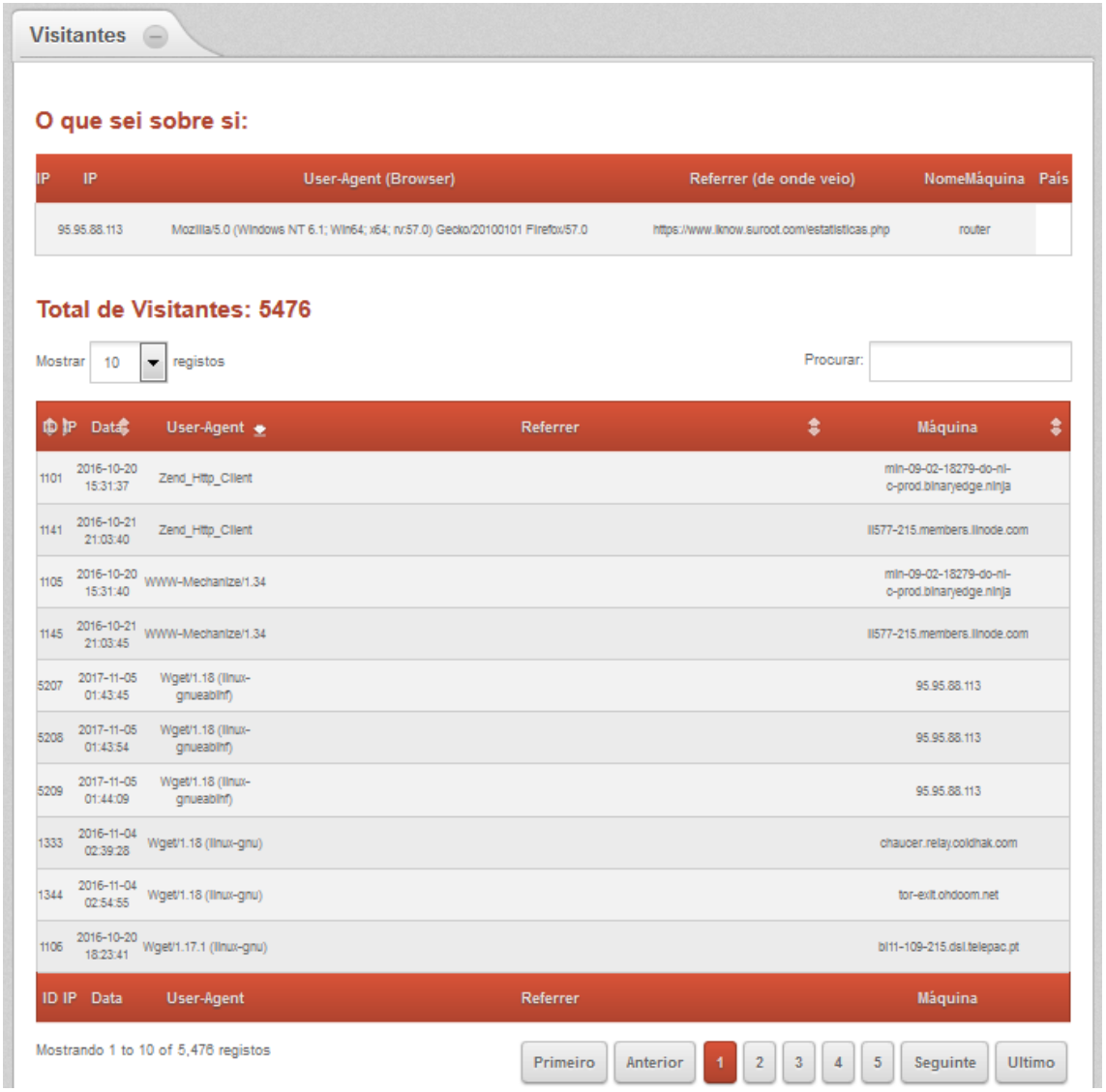


Figura 103 - Gráficos e Estatísticas - Visitantes do site

Figura 104 - gráficos e estatísticas - top anos de visitas

Top 10 de ips:

IP	Quantidade
127.0.0.1	836
85.244.109.215	781
	279
79.169.21.33	245
67.141.121.98	180
193.47.185.124	180
137.103.163.175	180
86.144.240.89	179
85.246.39.136	120
24.34.24.199	106

Figura 105 - Gráficos e Estatísticas - top 10 de ips

Top 10 de Browsers mais utilizados:

User-agent	Quantidade
Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:50.0) Gecko/20100101 Firefox/50.0	635
Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	583
Mozilla/5.0 (Windows NT 10.0; WOW64; rv:50.0) Gecko/20100101 Firefox/50.0	184
Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)	179
Go-http-client/1.1	178
	171
Mozilla/5.0 (Windows NT 6.1; rv:31.0) Gecko/20100101 Firefox/31.0	159
Mozilla/5.0 (Windows NT 5.1; rv:9.0.1) Gecko/20100101 Firefox/9.0.1	139
Python-urllib/3.5	120
Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2228.0 Safari/537.36	110

Figura 106 - Gráficos e Estatísticas - top 10 de browsers

Top 10 referrers:

referrer	Quantidade
	3802
http://376oeh24mdijht2e.onion/	638
https://gualdimpais.dtdns.net/estatisticas.php	122
http://79.169.21.33:80/	65
https://79.169.21.33:443/	62
https://gualdimpais.dtdns.net/tickets.php	49
https://95.95.88.113:443/	41
https://gualdimpais.dtdns.net/maquinas-temp.php	40
https://gualdimpais.dtdns.net/index.php	38
http://89.153.130.143:80/	37

Figura 107 - Gráficos e Estatísticas - top 10 de referrers

6.6 Código

Encontramos aqui o código utilizado para cada um dos itens do projecto. Códigos em HTML, PHP, MySQL, Python e Bash Shell, assim como fotografias das conexões em *hardware* para aceder as luzes LED utilizadas para dar sinal quando é detectado algum item procurado.

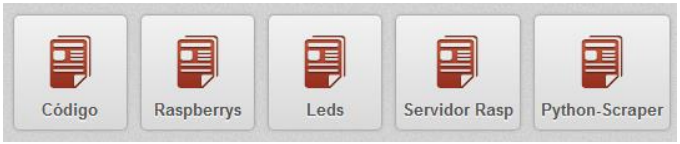


Figura 103- Código - Menu

Seguindo o código é possível colocar qualquer dos itens acima em funcionamento.

Código 1 - shell - Raspberry - Instalar bibliotecas necessárias

Código para raspberrys

Código para instalar as bibliotecas necessárias

instalar_python-mysqldb.sh

```
01. sudo apt-get update -y && sudo apt-get upgrade -y
02. sudo apt-get dist-upgrade -y
03. sudo apt-get install python-mysqldb -y
```

Código 2 - Raspberry - Registar inicio de fun. de equipamento

Código para raspberrys

Código para registar o inicio do funcionamento do equipamento, seu ip, nome, etc

pi_registar_maquina_servidor.py

copy text pop-up

```
01. #!/usr/bin/python
02. import MySQLdb
03.
04. host="192.168.1.201"
05. nome="p.."
06. passe="raspbe..2"
07. bd="tese"
08. # Open database connection
09. db = MySQLdb.connect(host,nome,passe,bd )
10.
11. # prepare a cursor object using cursor() method
12. cursor = db.cursor()
13.
14. #OBTER O IP PELO GOOGLE
15. #s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
16. #s.connect(("gmail.com",80))
17. #ip=(s.getsockname()[0])
```

Figura 108 - Código – Raspberrys - servidor

Há também indicação dos requisitos para o projecto. Neste caso em termos de *hardware* para que seja possível replicar as imagens fornecidas e ter o equipamento a mudar de cor conforme esteja a funcionar e/ou tenha descoberto algo.

Materiais - Raspberry / LED's

Materiais necessários

```
01. Breadboard: 1
02. LED vermelho: 1
03. LED verde: 1
04. Fios macho-fêmea: 3 (2 amarelos e 1 vermelho)
```

Figura 109 - Código - Requisitos para LED's

6.7 Máquinas – Disponibilidade e monitorização

Aqui pretende-se responder a algumas perguntas que são úteis quando o número de pesquisas é alto e para monitorização permanente do funcionamento: Quantas máquinas estão ligadas neste momento, qual a data da última transmissão, quais máquinas que mais trabalham e qual o volume de comunicações por mês. Clicando em cima da imagem da máquina podemos ver pormenores da mesma. Colocando o rato por cima do gráfico, temos um "balão" que nos informa do tempo que máquina esteve disponível.

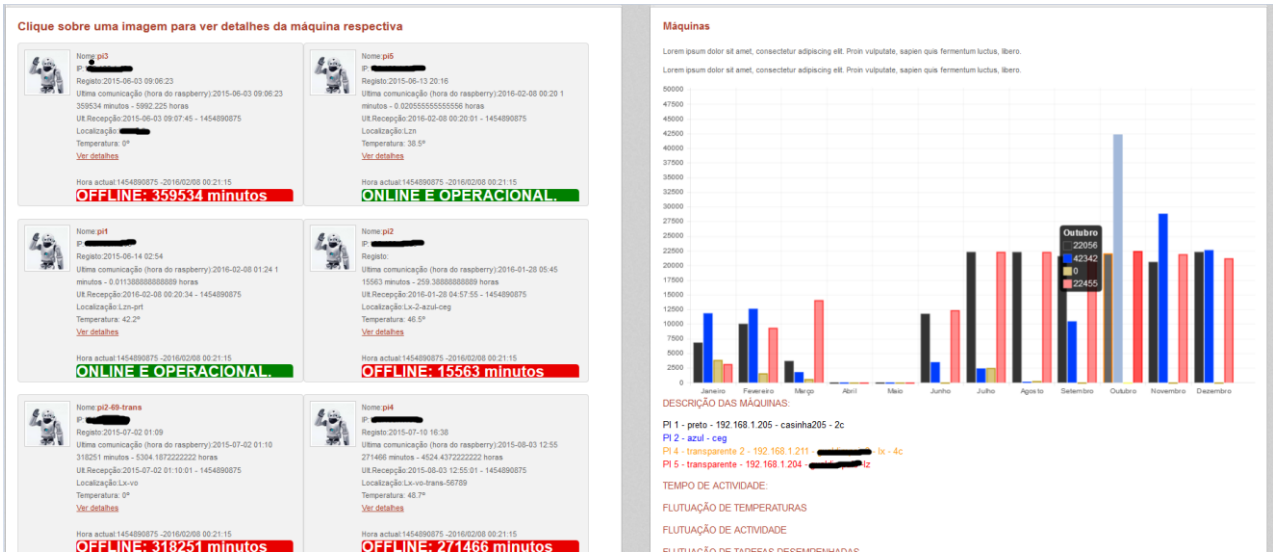


Figura 110 - Máquinas - Listagem de máquinas e sua Disponibilidade

Cada imagem pode ser clicada para aceder a mais informações. No geral, podemos ver há quanto tempo a máquina não está a responder, ou se a mesma está a funcionar. Localização da máquina, ip, hora da própria máquina e algo realmente importante: a temperatura (temperaturas muito elevadas podem indicar problemas futuros).

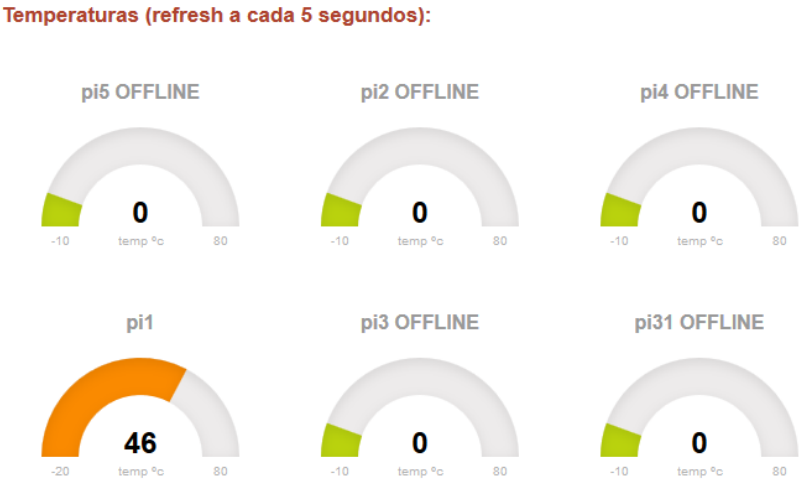


Figura 111 - Monitorização de temperaturas

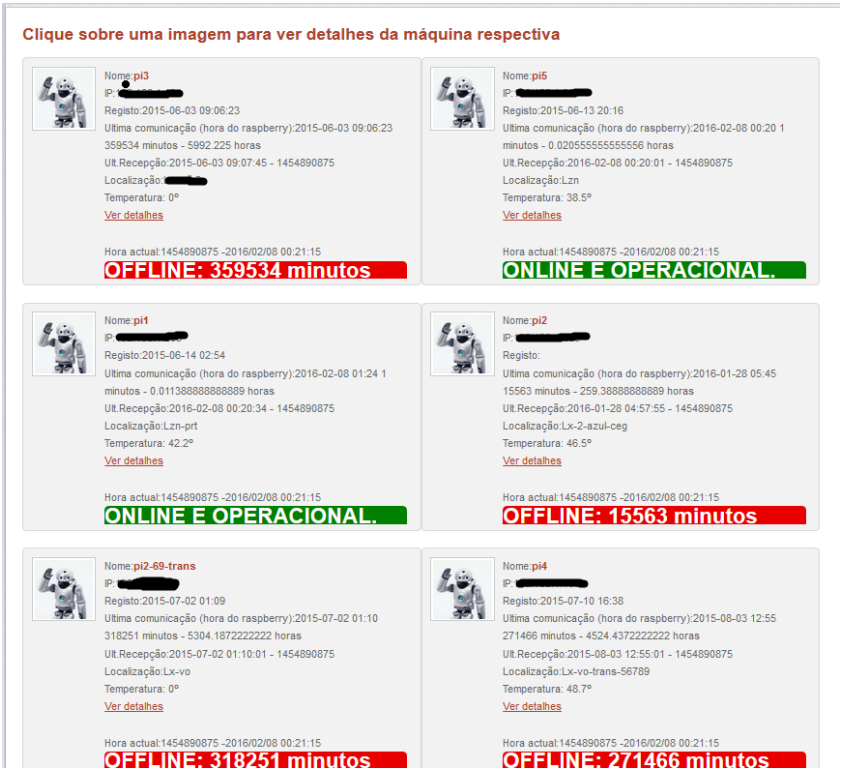


Figura 112 - Máquinas - Listagem de máquinas e seus pormenores

Período de actividade das máquinas. Por defeito, todas elas comunicam com o servidor de x em x tempo (definido). O gráfico mostra o tempo em minutos de actividade em cada mês. Maior barra significa maior disponibilidade. Cada cor é um equipamento diferente.

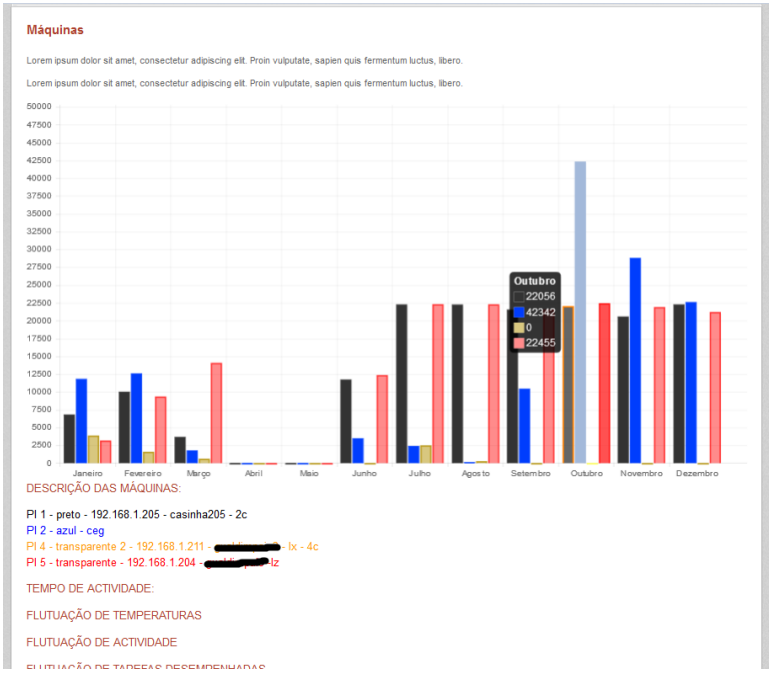


Figura 113 - Máquinas - Listagem de Disponibilidade

Cada cor é uma máquina, com localização diferente, temperatura, etc. Aqui podemos ver a disponibilidade em minutos do pi2, em cada mês, sendo 1 Janeiro, 2 Fevereiro, ...

pi2	5	0 Minutos 0 dias
pi2	6	3506 Minutos 4.8694444444444 dias
pi2	7	2429 Minutos 3.3736111111111 dias
pi2	8	160 Minutos 0.22222222222222 dias
pi2	9	10474 Minutos 14.547222222222 dias
pi2	10	42342 Minutos 58.808333333333 dias
pi2	11	28831 Minutos 40.043055555556 dias
pi2	12	22628 Minutos 31.427777777778 dias

Figura 114 - Máquinas - Tempo de Operacionalidade

Clicando em cima da imagem da máquina na folha principal, acedemos a mais pormenores da máquina.

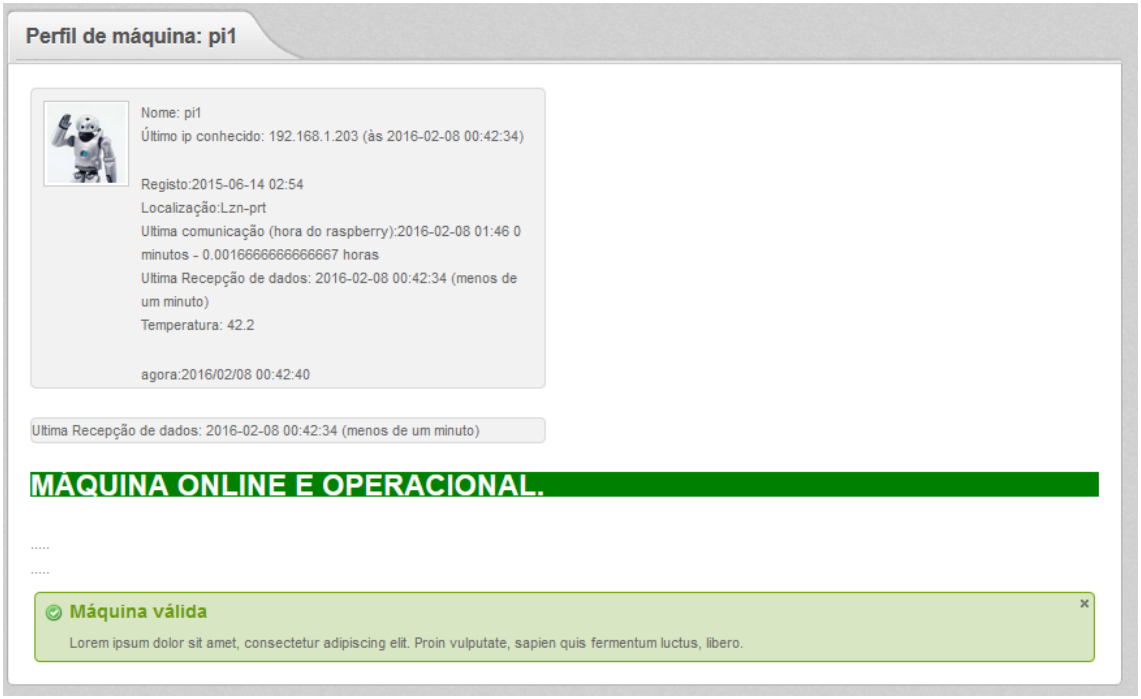


Figura 115 - Máquinas - Perfil de máquina

6.8 Área pessoal – chaves públicas e privadas, envio de informação

Aqui pode ser definida a chave pública, um calendário com as operações marcadas, notas pessoais e *tickets* com problemas.

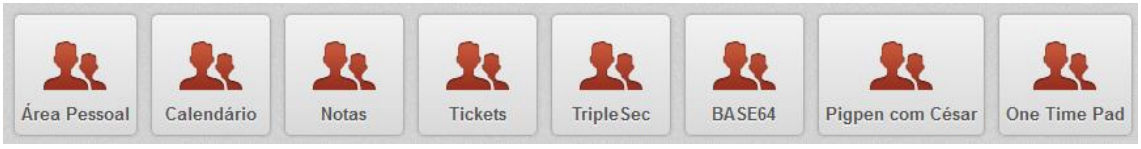


Figura 116 - Área Pessoal - Menu

A área pessoal pode ser actualizada com a chave pública para envio dos resultados. Caso esta chave não exista, o *mail* é enviado sem qualquer cifra. Esta secção, futuramente, apenas vai mostrar pormenores básicos dos utilizadores. Apenas um administrador poderá ver tudo. A

página mostra também alguns dados sobre o utilizador, à semelhança da página de autenticação e uma caixa mostrando se o utilizador é "válido" ou não (caso tenha sido bloqueado).

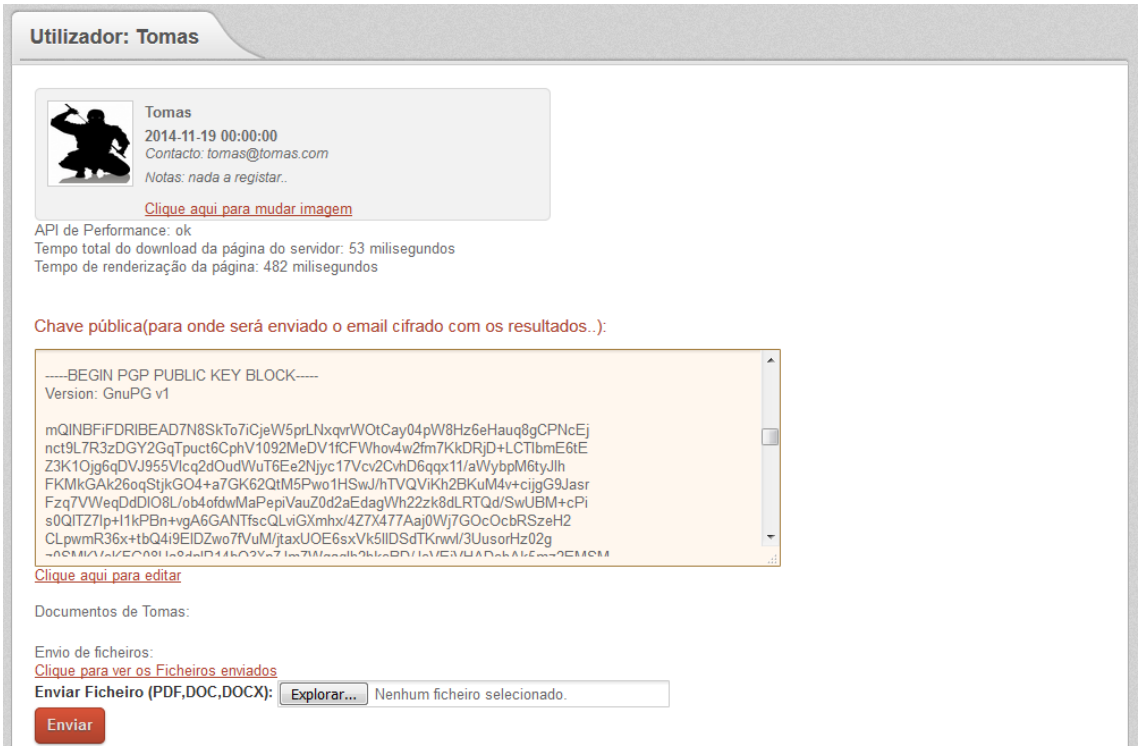


Figura 117 - Área Pessoal – Utilizador

Figura 118 - Área pessoal - Base64 e outras utilidades

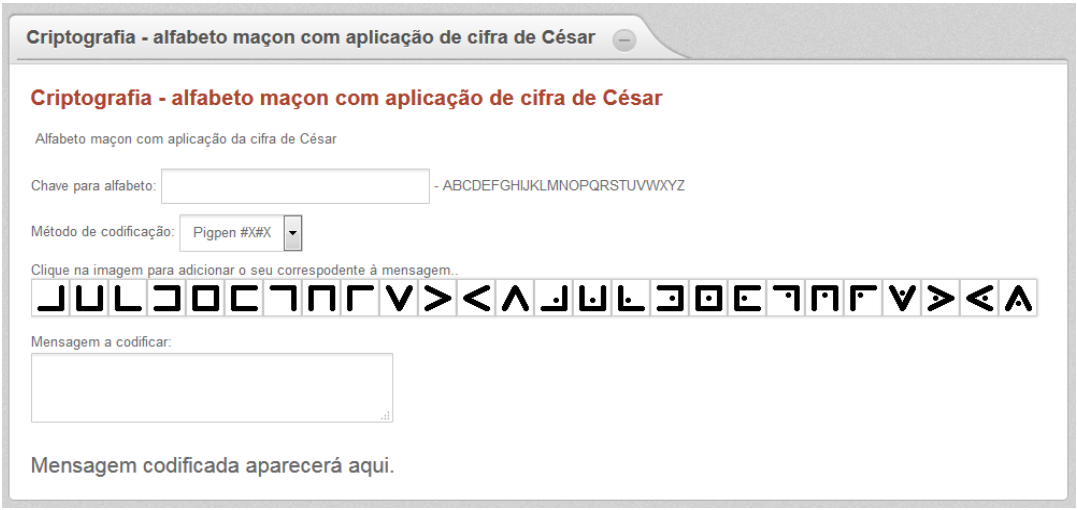


Figura 119 - Área Pessoal - Alfabeto maçon com aplicação da cifra de César

6.8.1 Área pessoal – registo de actividade, cifra e tickets

Notas e gráfico com actividade.

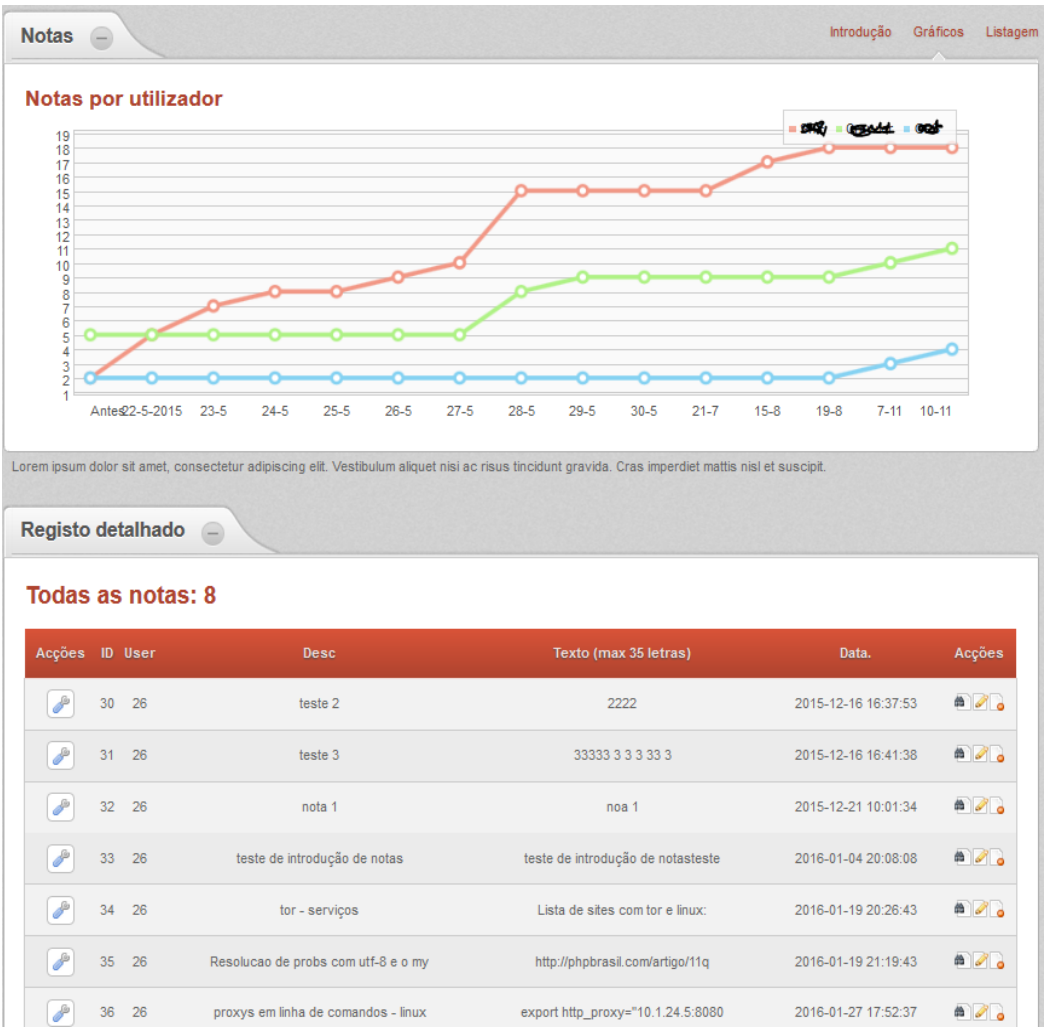


Figura 120 - Área Pessoal – Notas

Todas as notas são cifradas e armazenadas no servidor. É o cliente que faz a cifra e a decifra dos dados.

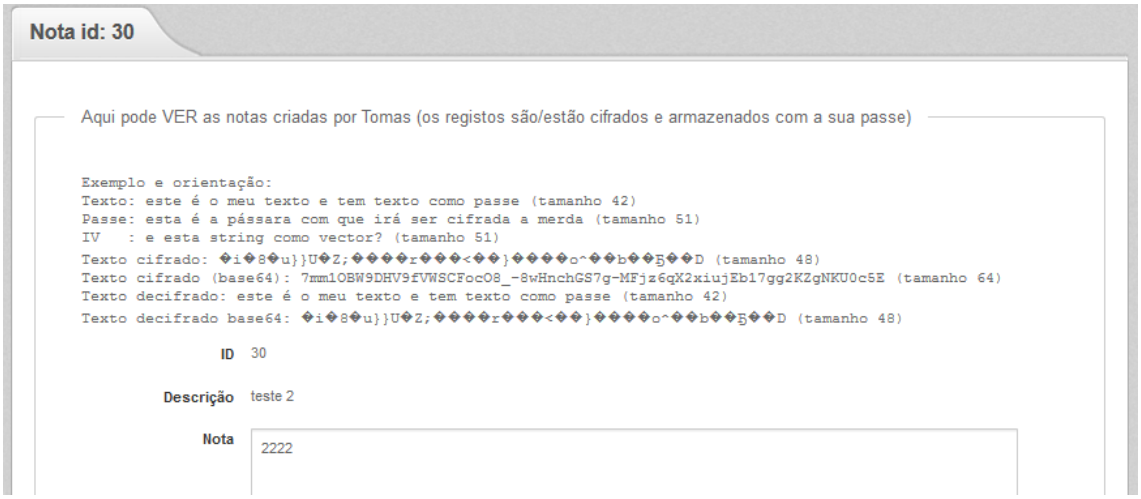


Figura 121 - Área Pessoal - Notas cifradas - Vista Geral

Os dados são cifrados uma primeira vez com o algoritmo *Blowfish*.

Figura 122 - Área Pessoal - Notas cifradas com cifra simétrica Blowfish

E depois cifrados em Base64 para poderem ser armazenados na base de dados, copiados, etc.

Figura 123 - Área Pessoal - Notas cifradas com simples Base64

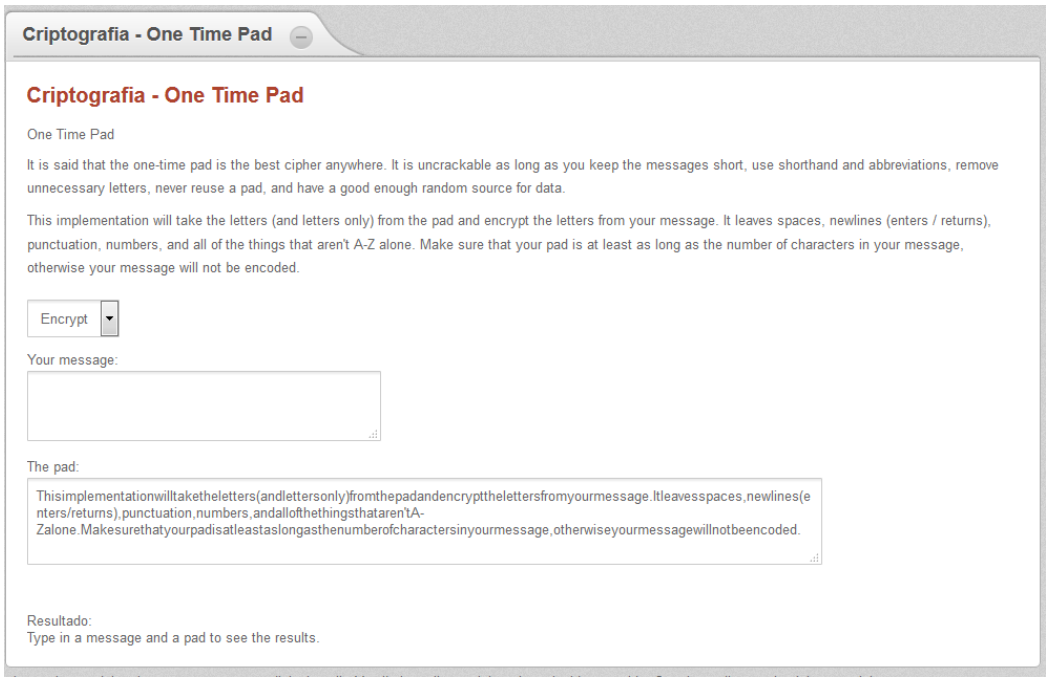


Figura 124 - Área Pessoal - Cifrar com One Time Pad (o único algoritmo que se bem utilizado é 100% seguro)

Os *tickets* permitem ao utilizador neste momento, criar notas pessoais. Futuramente apenas poderão criar e alterar, sendo que apenas os administradores verão e resolverão estes *tickets*, que terão dificuldades/sugestões e/ou outros problemas.

Permite a qualquer utilizador e/ou administrador, rapidamente criar, apagar e editar *tickets*, assim como atribuir uma classificação de gravidade.

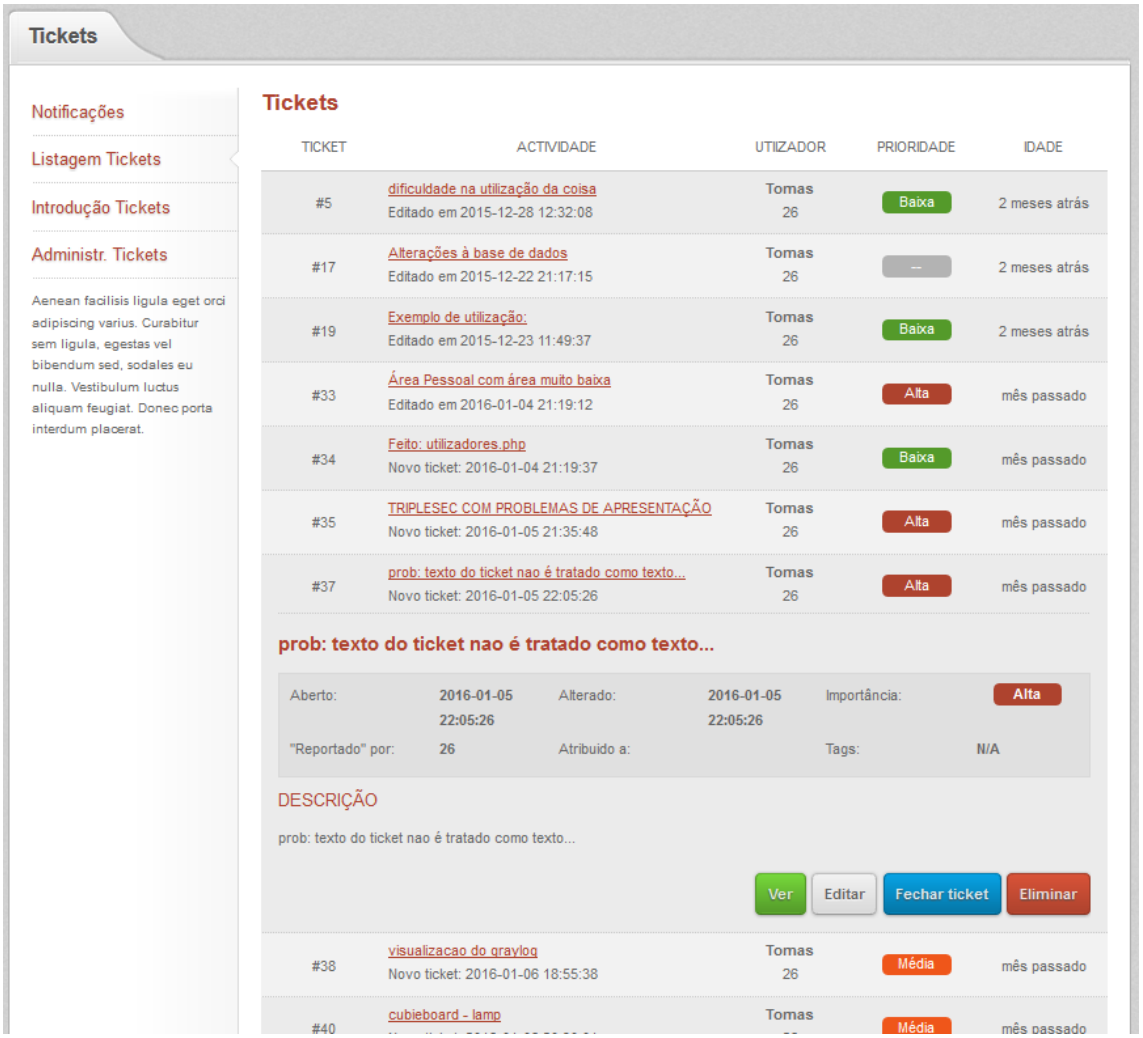


Figura 125 - Área Pessoal - Tickets

6.9 Registos/*logs* completos de acesso ao servidor

Registo detalhado de todos os eventos. *Login* errado ou correcto, inserção ou edição de dados, registo de novas máquinas, cliques, *user-agents*, de onde vieram os visitantes, entre outros, ...

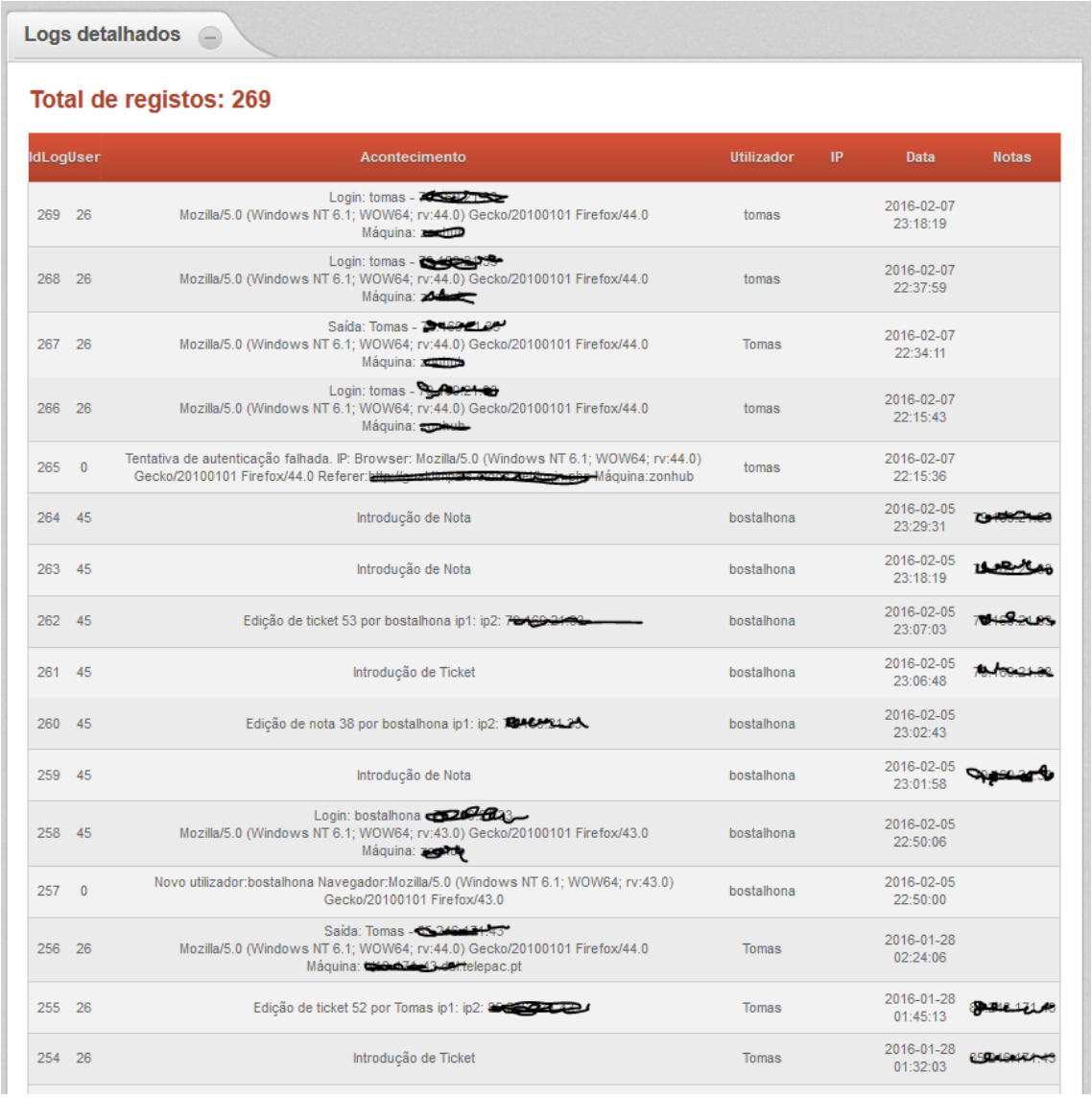


Figura 126 - Logs - Listagem de todos os eventos

7. Instação, código e configurações

7.1 Configuração de *firewall*

Série de comandos executados na CLI/linha de comandos/interface. Resumidamente, abrem todos os portos que precisamos ter abertos ao mundo, para que a aplicação possa funcionar sem problemas. Serviços 22-ssh, 80-web, 443- https, 3306-mysql, 56789-outro ssh).

```
# echo "instalar firewall e permissoes de acesso";
sudo apt-get install -y ufw;
sudo ufw allow 22;
sudo ufw allow 80;
sudo ufw allow 443;
```

```
sudo ufw allow 3306;
sudo ufw allow 56789;
sudo ufw enable;
echo "verficar";
netstat -tlpn;
```

7.2 Configurações de servidor *web Nginx*

Ficheiro de configuração do servidor web NginX. Escolhido por ser mais leve em termos de recursos do que o excelente Apache. A configuração mostra o seguinte: portos utilizados, obrigatoriedade do uso de https (comunicações seguras), localização dos ficheiros web, chaves utilizadas na cifra do servidor web, sua força criptográfica, entre outros.

```
server {
listen 443 ssl http2 default_server;
listen [::]:443 ssl http2 default_server;

root /var/www/html;
index index.php index.htm index.html;
server_name localhost;

location / {
try_files $uri $uri/ =404;
}

ssl_certificate /etc/nginx/ssl/nginx.crt;
ssl_certificate_key /etc/nginx/ssl/nginx.key;
ssl_protocols TLSv1 TLSv1.1 TLSv1.2;
ssl_prefer_server_ciphers on;
ssl_ciphers
EECDH+CHACHA20:EECDH+AES128:RSA+AES128:EECDH+AES256:RSA+AES256:EECDH+3DES:RSA+3DES:!
MD5;
ssl_dhparam /etc/nginx/ssl/dhparam.pem;
ssl_session_cache shared:SSL:20m;
ssl_session_timeout 180m;
resolver 8.8.8.8 8.8.4.4;
add_header Strict-Transport-Security "max-age=31536000";
#includeSubDomains" always;

location ~ \.php$
{
include snippets/fastcgi-php.conf;
fastcgi_pass unix:/run/php/php7.3-fpm.sock;
#fastcgi_pass unix:/run/php5-fpm.sock;
}

location ~ /\.ht
{
deny all;
```

```
}
}
```

7.3 Script automático de instalação – Servidor e plataforma

O *script* abaixo pode ser colocado num ficheiro e executado na linha de comandos. Resumidamente, faz a instalação completa de todos os pacotes necessários ao funcionamento do servidor Iknow. Instala servidor web, serviço ssh, servidor https, com geração de certificados no momento, *script* servidor *PHP7*, criação de utilizadores, entre outros. Deve ser sempre dada a prioridade ao script colocado no *GitHub*.

```
echo "----------------------------nginx-------------------------";
apt install nginx -y;
netstat -tlpn;
systemctl start nginx.service;
systemctl enable nginx.service;
apt install php7.0 php7.0-fpm php7.0-mysql php7.0-mbstring -y;
systemctl start php7.3-fpm.service;
service php7.0-fpm status;
echo "          ---------certificados";
sudo mkdir /etc/nginx/ssl;
sudo openssl req -x509 -nodes -days 365 -newkey rsa:2048 -keyout /etc/nginx/ssl/nginx.key -out
/etc/nginx/ssl/nginx.crt;
ls /etc/nginx/ssl/;
sudo openssl dhparam -out /etc/nginx/ssl/dhparam.pem 2048;
echo "         //---------//certificados";
cp ./default /etc/nginx/sites-available/default;
echo "//----------------------------//nginx-------------------------";

#echo "----------------------------ssh------------------------- ";
#echo "Copiar 'AllowUsers cripto' para etc/ssh/sshd_config";
#echo "Servidor ssh a correr no porto 56789";
#sed -i 's/#Port 22/Port 56789/' /etc/ssh/sshd_config;
#service ssh restart ;
#echo "//----------------------------//ssh-------------------------";

echo "---------------------------------- mariadb ----------------";
echo "instalar mariadb e criar utilizador - passe eh passe";
apt-get install mariadb-server mariadb-client -y;
echo "devia ser agora 'mysql_secure_installation;' mas para bem da instalacao do mysql/mariadb tem de
ser feito depois..";
service mysql start;
service mysql status;
echo "vamos criar o utilizador cripto com passe. Edite o ficheiro para mudar a passe";
sudo mysql -uroot -e "CREATE USER 'cripto'@'%' IDENTIFIED BY 'passe';";
sudo mysql -uroot -e "GRANT ALL ON *.* TO 'cripto'@'%' WITH GRANT option;";
sudo mysql -uroot -e "flush privileges;";
```

```
#PARA MUDAR A PASSE:
#SET PASSWORD FOR 'cripto'@'%' = PASSWORD('nova_passe');
echo "ver privilegios: SELECT User,Host,Password FROM mysql.user;";
sudo mysql -ucripto -p -e 'SELECT User,Host,Password FROM mysql.user;';
sudo mysql -uroot -p -e "create database tese;";
echo "importar bd";
#sudo mysql -uroot tese < tese.sql;
echo "#permitir acessos remotos ao porto e serviço 3306. Comentar bind-address";
#sed -i '/bind-address/c\#bind-address 192.168.0.1' /etc/mysql/my.cnf;
echo "se o servidor mariadb estiver num debian/ubuntu o ficheiro está em /etc/mysql/mariadb.conf.d/50-server.cnf";
sed -i '/bind-address/c\#bind-address 192.168.0.1' /etc/mysql/mariadb.conf.d/50-server.cnf
echo "------------------------ //mariadb --------------------------";


#---------------------------------- PhpMyAdmin -ficheiro my.sh-----------------------------
read -p "Instalar PhpMyAdmin? <y/N> " prompt;
if [ "$prompt" = "y" ];
then apt-get install -y phpmyadmin;
ln -s /usr/share/phpmyadmin /var/www/html/;
echo "http://192.168.1.200/phpmyadmin para entrar no PhpMyAdmin";
fi
#---------------------------------- // PhpMyAdmin --------------------------

echo "#outros ficheiros de autenticacao obrigatorios";
echo "copiar ficheiros para a pasta certa e dar permissoes";
mv testar_autenticacao3.php /home/cripto/;
sudo chown www-data:www-data /home/cripto/testar_autenticacao3.php;
mv paxs1.php /home/cripto/;
sudo chown www-data:www-data /home/cripto/paxs1.php;

sudo usermod -a -G www-data cripto; id cripto;
#sudo reboot;

# ---------------------------------- data por NTP ------------------
echo "Sincronizar data com ntp portugues";
apt install ntpdate -y;
echo "ou em caso de uso de systemd:";
echo "NTP=0.europe.pool.ntp.org" >> /etc/systemd/timesyncd.conf;
echo "FallbackNTP=3.europe.pool.ntp.org" >> /etc/systemd/timesyncd.conf;
timedatectl set-ntp true;
timedatectl status;
```

7.4 Configurações de acesso – *critico*!

Ficheiro dados.py - Neste ficheiro ficam as configurações de acesso e/da base de dados. O *host* pode ser um endereço ip ou domínio. Nome e passe, devem ser obtidos do servidor iKNOW, quer manualmente pedindo, ou manualmente registando e obtendo de imediato estas

credenciais. Pode estar aqui uma falha de segurança, pois o cliente pode criar um DOS/DDOS. Por esta razão, esta possibilidade foi restrita nas últimas versões.

#host="*192.168.1.201*"
host="*dominiodoservidor_ou_ip.*net"
nome="*pi*"
#nome eh o utilizador .... nome="pi5"
passe="...."
bd="*tese*"
loc="*LZ-wca-vo*"
maquina="*pi5*"

7.5 Código do *bot Telegram*

```
#!c:/Python37/python
# -*- coding: utf-8 -*-

import re
from datetime import date
import datetime
import time
import mysql.connector
import hashlib
import telebot
import requests
import io
from newspaper import Article

utilizador=username="tomas"
entidade="cncs"
ficheiro="C:/python37/artigo_utf8.htm"
Palavras_Chave=""

print(" ---------- Inicio em 25 segundos %s----------" % datetime.datetime.now().time())

resultado=decoded.decode("utf-8")
#print(resultado)
""" FIM BASE 64 """


def obter_comando(url_noticia):
        comando="C:/Python37/python.exe/osints_obter_noticia_telegrama.py %s" % url_noticia
        return comando;

#-------------------------------------------------------------------------------
def obter_pagina_web_do_telegrama(html):
        #from newspaper import Article
        a = Article(html)
        a.download()
        a.parse()

        titulo = a.title
        titulo= re.sub('\"',"'", titulo.rstrip())
        texto = re.sub('\"',"'", texto.rstrip())

def obter_source(url):
```

```
        #import requests
        html = requests.get(url).content
        #html = html.decode("utf-8")
        #soup = BeautifulSoup(html)
        return html

def obter_registos_da_bd_telegrama(url):
        char_de_asneira=")"
        ultimo_char=url[-1:]
        if (ultimo_char==char_de_asneira):
                #print(ultimo_char)
                #print("asneira")
                url=url[:-1]
        link=fonte=url

        a = Article(url, language='en') # isto eh importante..

        a.download()
        a.parse()

        codigo_html=a.html
        resumo= a.text[:150]
        titulo=a.title
        texto=a.text
        imagem=a.top_image
        autor=a.authors

        ############### hashes e data ###############
        hoje=date.today()
        dia=hoje.day
        mes=hoje.month

        ano=hoje.year
        semana=datetime.date(ano, mes, dia).isocalendar()[1]
        imagem=a.top_image
        noticia=texto
        resumo=texto[0:100]
        hora=time.strftime("%H:%M:%S")

        import hashlib
        #mystring = input('Enter String to hash: ')
        hash_object = hashlib.sha256(codigo_html.encode())
        print(hash_object.hexdigest())
        ############### //hashes e data ###############

        ############## base64 do codigo ##################
        #print(codigo_html)
        codigo_base64 = base64.b64encode(codigo_html.encode())
        ############## // base64 do codigo ##################

        data = a.publish_date
        data1=obter_data()

        if data==None:
                data="%s-%s-%s" % (ano,mes,dia)
                dia=data1[0]
                mes=data1[1]
                ano=data1[2]
        if data=="":
                data="%s-%s-%s" % (ano,mes,dia)
                dia=data1[0]
                mes=data1[1]
                ano=data1[2]
```

```
		titulo= re.sub('\"',"'", titulo.rstrip())
		texto = re.sub('\"',"'", texto.rstrip())

		#print("Codigo:\n%s" % codigo_html)
		#print("Texto:\n%s" % texto)
		#print("Titulo:%s" % titulo)
		print("Data:%s" % data)
		print("Autor(es):%s" % autor)
		hash=""
		#print("Hash:%s" % hash)


		#Palavras_Chave=""
		categoria="telegrama"
		#ficheiro="C:/python37/artigo_utf8.htm"

		with io.open(ficheiro, "w", encoding="utf-8") as f:
			CONTEUDO_FICHEIRO=('Titulo:%s<br/>Autor:%s<br/>Origem:%s<br/>Imagem:%s<br/><img
src="%s"><br/>Data:%s<br/>Palavras_Chave:%s<br/>Texto:%s<br/><br/>Codigo:%s<br/>' %
(titulo,autor,url,imagem,imagem,data,Palavras_Chave,texto,codigo_html))
			f.write(CONTEUDO_FICHEIRO)

		try:
			mydb= mysql.connector.connect(host='localhost',		database='logins', user='root',
password='')
			mycursor = mydb.cursor()


			#codigo_base64=semana=ano=mes=dia=hora=""
			fonte=link=url
			categoria="telegrama"
			entidade="cncs"
			#sql = """INSERT INTO osints(titulo, texto, data, semana, link,
fonte,imagem,ficheiro,categoria,resumo,hash,codigo_base64,ano,mes,dia,entidade,noticia,hora,username) VALUES
(%s,%s,%s,%s,%s,%s,%s,%s, %s, %s, %s, %s, %s, %s, %s, %s,%s,%s,%s,%s,%s)"""
			sql = """INSERT INTO osints(titulo, texto,
data,semana,link,fonte,imagem,categoria,codigo,codigo_base64,ano,mes,dia,entidade,hash) VALUES
(%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s)"""

			val = (titulo, texto,
data,semana,link,fonte,imagem,"telegrama",codigo_html,codigo_base64,ano,mes,dia,entidade,hash)
			#, semana, link,
fonte,imagem,ficheiro,categoria,resumo,hash_pagina,codigo_base64,ano,mes,dia,entidade,noticia,hora,username)
			mycursor.execute(sql, val)

			mydb.commit()
			print(mycursor.rowcount, "Registo inserido. Posição ", mycursor.lastrowid)

		except mysql.connector.Error as err:
			print("erro ao conectar bd:", err)


def obter_registos_da_bd_telegrama2(url):
		ficheiro="C:/python37/artigo_utf8.htm"

		char_de_asneira=")"
		ultimo_char=url[-1:]
		if (ultimo_char==char_de_asneira):
			#print(ultimo_char)
			#print("asneira")
			url=url[:-1]
		link=fonte=url
```

```
	#a = Article(url)
	a = Article(url, language='en')
	a.download()
	a.parse()

	titulo = a.title
	texto = a.text

	titulo= re.sub('\"',"'", titulo.rstrip())
	texto= re.sub('\"',"'", texto.rstrip())


	html=obter_source(url)
	#html_utf8=html.decode("utf-8")
	html_utf8=html.encode('utf-8')

	#print(html_utf8) #debug apenas
	#import base64
	b64_cifrado=base64.b64encode(html)
	b64_utf=b64_cifrado.decode("utf-8")
	codigo=b64_utf
	codigo_base64=b64_cifrado

	hoje=date.today()
	dia=hoje.day
	mes=hoje.month
	ano=hoje.year
	semana=datetime.date(ano, mes, dia).isocalendar()[1]
	imagem=a.top_image
	noticia=texto
	resumo=texto[0:100]
	hora=time.strftime("%H:%M:%S")
	hash=hashlib.sha256(html_utf8.encode('utf-8')).hexdigest()

	data = a.publish_date
	data1=obter_data()

	if data==None:
		data="%s-%s-%s" % (ano,mes,dia)
		dia=data1[0]
		mes=data1[1]
		ano=data1[2]
	if data=="":
		data="%s-%s-%s" % (ano,mes,dia)
		dia=data1[0]
		mes=data1[1]
		ano=data1[2]

	print("titulo:\t%s" % titulo)
	print("data:\t%s" % data)
	#print("texto:\n%s" % texto)
	#print("html utf8:\n%s" % html_utf8)
	#print("codigo:\n%s" % codigo) #cifrado b64
	print("resumo:\t%s" % resumo)
	print("imagem:\t%s" % imagem)
	print("hash:\t%s" % hash)


	#import io
	#with io.open("C:/xampp/htdocs/artigo_utf8.txt", "w", encoding="utf-8") as f:
	#FICHEIRO1="C:/Users/helder.tomas/Downloads/xamppx64-7.3.5-1-VC15/htdocs/a-
logins/artigo_utf8.txt"
```

```
		Palavras_Chave=""
		categoria="telegrama"
		with io.open(ficheiro, "w", encoding="utf-8") as f:
			CONTEUDO_FICHEIRO=('Titulo:%s<br/>Autor:%s<br/>Origem:%s<br/>Imagem:%s<br/><img
src="%s"><br/>Data:%s<br/>Palavras_Chave:%s<br/>Texto:%s<br/><br/>Codigo:%s<br/>' %
(titulo,a.authors,url,a.top_image,a.top_image,data,Palavras_Chave,texto,codigo))
			f.write(CONTEUDO_FICHEIRO)


		#tudo funciona. Enviar para BD mesmo sem categoria CONHECIDA
		#import mysql.connector
		try:
			mydb= mysql.connector.connect(host='localhost',		database='logins', user='root',
password='')
			mycursor = mydb.cursor()

			#sql = """INSERT INTO osints(titulo, texto, data, semana, link,
fonte,imagem,ficheiro,categoria,resumo,hash,codigo_base64,ano,mes,dia,entidade,noticia,hora,username) VALUES
(%s,%s,%s,%s,%s,%s,%s,%s, %s, %s, %s, %s, %s, %s, %s, %s,%s,%s,%s,%s)"""
			sql = """INSERT INTO osints(titulo, texto,
data,semana,link,fonte,imagem,categoria,codigo_base64,ano,mes,dia,entidade,hash) VALUES
(%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s,%s)"""

			val = (titulo, texto,
data,semana,link,fonte,imagem,"telegrama",codigo_base64,ano,mes,dia,entidade,hash)
			#, semana, link,
fonte,imagem,ficheiro,categoria,resumo,hash_pagina,codigo_base64,ano,mes,dia,entidade,noticia,hora,username)
			mycursor.execute(sql, val)

			mydb.commit()
			print(mycursor.rowcount, "Registo inserido. Posição ", mycursor.lastrowid)

		except mysql.connector.Error as err:
			print("erro ao conectar bd:", err)
		return None
#-------------------------------------------------------------------------------


def obter_data():
		#from datetime import date
		hoje=date.today()
		dia=hoje.day
		mes=hoje.month
		ano=hoje.year
		#print ("dia:%s mes:%s ano:%s" % (dia,mes,ano))

		#import datetime
		semana=datetime.date(ano, mes, dia).isocalendar()[1]
		#print("Semana do ano:",semana)

		#import time
		hora=time.strftime("%H:%M:%S")
		#print(hora)

		#agora=now.strftime()
		dia=time.strftime("%d")
		mes=time.strftime("%m")
		return dia,mes,ano,semana,hora

def mysql_inserir(dia,mes,ano,semana,link,noticia,hora,utilizador,hash):
		#import mysql.connector
		try:
			mydb= mysql.connector.connect(host='localhost',		database='logins',  user='root',
```

```
password='')
                        mycursor = mydb.cursor()

                        sql = 'INSERT INTO telegrama(dia,mes,ano,semana,link,noticia,hora,username,hash) VALUES
(%s, %s, %s, %s, %s, %s, %s, %s, %s)'
                        val = (dia,mes,ano,semana,link,noticia,hora,username,hash)
                        mycursor.execute(sql, val)

                        mydb.commit()
                        print(mycursor.rowcount, "Registo inserido. Posição ", mycursor.lastrowid)

        except mysql.connector.Error as err:
                        print("erro ao conectar bd:", err)
                        return None

def hash_noticia_sha256(mensagem):
        #import hashlib
        noticia = hashlib.sha256(mensagem.encode('utf-8')).hexdigest()
        print("\tsha256 da noticia: ", noticia)
        #hashes existente: https://www.pythoncentral.io/hashing-strings-with-python/
        return noticia


#@bot.message_handler(func=lambda m: True)
#def echo_all(message):
#       print(message.text)
#       #bot.reply_to(message, "enviado")

#@bot.message_handler(regexp='((http?):(((//)|(\\\\))+([\w\d:#@%/;$()~_?\+-=\\\.&](#!)?)*)')
#def echo_all(message):
#       link = re.search('((https?):(((//)|(\\\\))+([\w\d:#@%/;$()~_?\+-=\\\.&](#!)?)*)', message.text)
#       print(message.text)
#       print("MENSAGEM COMPLETA: " + message.text)
#       print("SOMENTE O LINK DA MENSAGEM: " + link.group(0))


#FAZ TUDO
#x=obter_registos_da_bd_telegrama("https://www.hackingarticles.in/linux-for-pentester-pip-privilege-
escalation/")

def obter_tudo(url):
        #a = Article(url)
        a = Article(url, language='en')
        a.download()
        a.parse()
        Palavras_Chave=""

        texto=a.text
        titulo=a.title
        titulo= re.sub('\"',"'", titulo.rstrip())
        texto = re.sub('\"',"'", texto.rstrip())

        data=""
        try:
                        data = a.publish_date
                        data = re.sub('\"',"'", data.rstrip())
        except:
                        pass

        html = requests.get(url).content
        #html_utf8=html.decode("utf-8")
        #html = re.sub('\"',"'", html.rstrip())
        print(html)
```

```
		b64_cifrado=base64.b64encode(html)
		b64_utf=b64_cifrado.decode("utf-8")
		codigo=b64_utf

		FICHEIRO1="C:/python37/artigo_utf8.htm"
		with io.open(FICHEIRO1, "w", encoding="utf-8") as f:
			CONTEUDO_FICHEIRO=('Titulo:%s<br/>Autor:%s<br/>Origem:%s<br/>Imagem:%s<br/><img
src="%s"><br/>Data:%s<br/>Palavras_Chave:%s<br/>Texto:%s<br/>Codigo:<br/>%s<br/>' %
(titulo,a.authors,url,a.top_image,a.top_image,data,Palavras_Chave,texto,codigo))

			f.write(CONTEUDO_FICHEIRO)

#obter_tudo("https://www.darkreading.com/risk/cybercriminals-target-budding-cannabis-retailers/d/d-
id/1335184")


#codigo original
import sched, time
s = sched.scheduler(time.time, time.sleep)
def do_something(sc):
		print ("%s ---------------------------- Polling de 25 segundos... ----------------------------" %
(datetime.datetime.now().time()))
		bot = telebot.TeleBot("COLOCAR_AQUI_A_API")

		@bot.message_handler(commands=['start', 'help'])
		def send_welcome(message):
			bot.reply_to(message, "Bom dia, tudo bem?")


		@bot.message_handler(func=lambda message: True)
		def echo_all(message):
			bot.reply_to(message, message.text)
			print("\n\n-------- Obtenção da informação via telegrama --------\n")

			try:
				link=re.findall(r'(https?://\S+)', message.text)
				link=link[0]
			except:
				link="sem link"

			print("\tLink(s)\t: %s" % link)

			Artigo=link
			obter_comando(Artigo)

			data=obter_data()
			dia=data[0]
			mes=data[1]
			ano=data[2]
			semana=data[3]
			hora=data[4]

			print("\tData\t: %s %s %s semana:%s hora:%s" % (dia,mes,ano,semana, hora))
			noticia=message.text
			data="%s-%s-%s" % (dia,mes,ano)
			print("\tNoticia\t: %s" % noticia)

			#hash
			hash=hash_noticia_sha256(noticia)

			#inserir apenas no fim e deve-se verificar primeiro se não existe o mesmo hash
			mysql_inserir(dia,mes,ano,semana,link,noticia,hora,utilizador,hash)
```

```
                    url=str(link)
                    print("\tUrl:\t%s" % url)

                    print("\n\n-------- Inserção e obtenção da informação na BD --------\n")
                    obter_registos_da_bd_telegrama(url)

        bot.polling()

        # do your stuff
        s.enter(25, 1, do_something, (sc,)) #1 sao as vezes?

s.enter(25, 1, do_something, (s,))
s.run()
```

7.6 Monitor sempre activo – Potenciais dashboards

#impedir que se desligue o monitor. Copiar para "" e colocar o codigo antes da linha "exit 0".
#fazer depois reboot
sudo nano /etc/rc.local
sh -c 'setterm -blank 0 -powersave off -powerdown 0 < /dev/console > /dev/console 2>&1'

7.7 *Tweaks* - Aumento de performance

```
echo "ssl_session_cache shared:SSL:5m;" >> /etc/nginx/nginx.conf;
echo "ssl_session_timeout 1h;" >> /etc/nginx/nginx.conf;
echo    "add_header    Strict-Transport-Security    "max-age=15768000;
includeSubDomains" always;" >> /etc/
```

7.8 Serviços *TOR*

Cada equipamento que se liga e trabalha com a plataforma iKNOW, utiliza serviços TOR. Isto é obrigatório na medida que precisamos que o equipamento se ligue à rede TOR se houver necessidade de baixar informações que eventualmente possam estar nessa rede. Ao mesmo tempo, está previsto que nas versões futuras do iKNOW, as comunicações entre o equipamento e o servidor se façam nesta rede. O objectivo para o futuro é que cada equipamento e utilizador estejam totalmente anonimizados de forma a utilizarem a plataforma sem qualquer receio de vigilância.

sudo apt-get update -y;
sudo apt-get install tor -y;
#Editar /etc/tor/torrc e acrescentar:
#HiddenServiceDir /var/lib/tor/other_hidden_service/
#HiddenServicePort 80 127.0.0.1:80
#HiddenServicePort 22 127.0.0.1:56789
echo "HiddenServiceDir /var/lib/tor/other_hidden_service/" >> /etc/tor/torrc;

```
echo "HiddenServicePort 80 127.0.0.1:80" >> /etc/tor/torrc;
echo "HiddenServicePort 22 127.0.0.1:22" >> /etc/tor/torrc;
echo "HiddenServicePort 443 127.0.0.1:443" >> /etc/tor/torrc;
echo "RunAsDaemon 1" >> /etc/tor/torrc;

export SERVICE_DIR=/var/lib/tor/other_hidden_service/;
mkdir $SERVICE_DIR;
chmod 700 $SERVICE_DIR;
chown debian-tor.debian-tor $SERVICE_DIR;
systemctl enable tor;
systemctl start tor;
cat $SERVICE_DIRhostname;
```

7.9 Configuração do balanceador *HAProxy*

O software HAProxy foi o escolhido para fazer o balanceamento dos serviços http/s e bases de dados MySQL/MariaDB, devido à sua facilidade de utilização, rápida configuração, mas principalmente por ser quase o standard em balanceadores.
O algoritmo escolhido foi *round robin*, de forma a que a carga fosse uniformemente distribuída. Foram utilizadas sessões e o modo "*sticky*" para que a comunicação entre os diversos servidores mysql e php se mantivesse igual, para evitar necessidade de novas autenticações (para cada pedido para servidor diferente).

7.10 *Cron* – Automatização

O cron é o serviço que é utilizado para executar os comandos automaticamente a data e hora fixa. É importante a sua presença para actualizar o sítio web e enviar/receber informações da página web para o servidor.

```
*/4 * * * * /home/cripto/pi_actualizar_servidor.py
*/10 * * * * cd /var/www/html/funcional-feeds_jor_nal_nego_cios/ && python3 ./osint_jornal_de_negocios.py
##*/5 * * * * /home/pi/actualizar.sh ##*/3 * * * * python3 /var/www/html/scrape10-eur.py
##funca mas enche muito o disco: 4gb
#*/59 * * * * cd /var/www/html/ && python3 ./scrape10-eur.py >> /var/www/html/scrape10.log # 2>&1

##Ultimo:
##*/59 * * * * cd /var/www/html/ && python3 ./scrape10-eur.py

##sabado 8:05 da manha
#5 15 * * Sat root /home/cripto/backup.sh

#todos os dias as 8 minutos
#00 09 * * * cd /var/www/html/ && python3 ./osint_imagens_dos_jornais_do_dia.py
```

```
1/25 09 * * * cd /var/www/html/ && python3 ./osint_imagens_dos_jornais_do_dia.py
```

7.11 Plataforma web – *HTML, javascript, PHP, Python, …*

A plataforma web assenta em muitas páginas com código *html*, intercalado com *scripting* em linguagem *javascript e php.*

Algumas das opções do site utilizam o PHP como forma de chamar directamente comandos Python no servidor, como por exemplo, baixar a notícia de um sítio web. O PHP é muito poderoso no sentido também da ponte entre o código, a base de dados e o interface web, na facilidade de programação e na rapidez com que processa.

Todo o *scrape* e *crawl* é feito via *scripts Python* em consola. O resultado é enviado para a base de dados. Posteriormente outros *scripts PHP* e *Python*, irão aceder à base de dados para obter tarefas (baixar páginas, enviar, processar, enviar relatórios, …)

7.11.1 Exemplo de código-fonte de *crawler PHP*, comentado

```
<?php
//inicio da sessão. Guarda-se na sessão se o utilizador está registado e inicializado, assim como algumas variáveis de ambiente.
session_start();
date_default_timezone_set("Europe/Lisbon"); //Localização do site em fuso horário. Isto é importante para não termos fusos horários diferentes. Todos os equipamentos também devem estar sincronizados com esta hora.

if(!isset($_SESSION['username']) || empty($_SESSION['username'])){
  header("location: login.php"); //se username não existir, utilizador não está autenticado e por isso é enviado para a página de login
  exit;
}

require('url_to_absolute/url_to_absolute.php'); //este ficheiro que é obrigatório, simplifica a divisão do caminho. Util para não misturar domínios.
?>
…
      Olá, <b><?php echo $_SESSION['username']; ?></b>. Bem-vindo.
…
<?php
            $base = "https://www.ecce.gov.pt/"; //iniciamos o crawler neste endereço para fins de teste e prova de conceito. Podia ser colocado um formulário a pedir ao utilizador, o site onde este queria começar e navegar/crawlar

function crawl_page($url, $depth = 10)
{
  static $seen = array();
  if (isset($seen[$url]) || $depth === 0) {
    return;
  }

  $seen[$url] = true;

  $dom = new DOMDocument('1.0');
  @$dom->loadHTMLFile($url);

  $anchors = $dom->getElementsByTagName('a');
  foreach ($anchors as $element) {
```

```
   $href = $element->getAttribute('href');
   if (0 !== strpos($href, 'http')) {
     /* this is where I changed hobodave's code */
      $host = "http://".parse_url($url,PHP_URL_HOST);
      $href = $host. '/' . ltrim($href, '/');
   }
   crawl_page($href, $depth - 1);
 }

 echo "<b>".$url ."</b><br/> ";
 echo htmlspecialchars($dom->saveHTML())."<br><br/><br/>";
 //echo PHP_EOL,"<br />","CONTENT:",PHP_EOL,htmlspecialchars($dom->saveHTML()),PHP_EOL,PHP_EOL,"  <br
/><br />";
}

crawl_page($base, 10);
?>

<br/> <br/> <br/> <br/> <br/>
</body>
</html>
```

7.12 iKNOW_bot – Bot do *Telegram*

Esta é das principais ferramentas
Foi criado para o iKNOW, uma aplicação que faz uso do software "Telegram"[237]. Esse automatistmo/*bot* iKNOW_bot do Telegrama permite a recepção de mensagens e a automatização da recolha de informações.

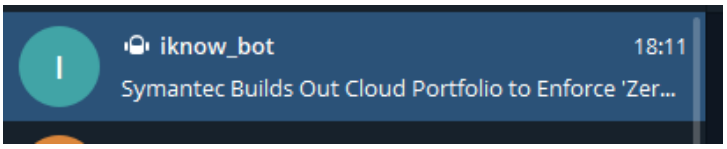


Figura 127 - Bot iknow_bot no telegrama

[237] https://telegram.org/, aplicação gratuita para telemóvel e computador, de mensagens e transferência de ficheiros e informações, de forma rápida, simples e segundo os próprios de forma cifrada e segura

8. Caso prático de OSINT

ÍNDICE

Introdução e resumo do problema

Foi obtido um pedido de ajuda relativamente a um IP que fazia *portscan* à nossa instituição X. O pedido apenas trazia o IP e evidências do *portscan* aos sites e serviços. O "relatório" abaixo mostra os procedimentos e *modus operandi* que levaram à identificação total do atacante, meramente recorrendo a OSINT, sem qualquer tentativa ou violação de barreiras técnicas nem interacção com o "potencial" atacante.
Nota: A numeração marca os passos efectuados ou a sequência de acontecimentos.

Acções tomadas

1. Foi feito pedido de ajuda por parte do responsável pela *firewall* a um IP que fez *portscan* à rede institucional):

"…
Foi identificado ontem (xx/xx/2019) na FW uma ação de portscan, por volta das 19:08.
Após consulta no abuseipdb.com o ip em questão não está reportado.
Podem validar nas feeds que têm acesso, se o IP público xxx.xx.xxx.153 está associado a alguma atividade maliciosa, sff.
…"

2. Departamento de Operações informa que não encontrou evidências nem "abusos".
3. Vamos analisar nós:
   - IP de origem brasileira faz um *portscan* massivo aos IPs públicos da entidade x (a nossa que não nos interessa identificar aqui).
   - Verificando pelo IP, verifica-se que o IP não é de VPN.
   - Ser um IP brasileiro é já uma boa pista já que da experiência pessoal, somos (Portugal) um bom alvo de testes por parte dos brasileiros.
   - Sem entrar em manobras ofensivas, testou-se usar o IP brasileiro atacante no browser (nota: foi usada a rede externa para aceder aquele IP).
   - Verificou-se que o IP tem um site a correr no porto padrão 80. Não foi feita mais nenhuma tentativa de acessos a serviços, já que tal situação podia ser considerada um ataque da nossa parte.
   - Está a correr um servidor web: http://xxx.xxx.xxx.xxx/login.php, simples, não usa certificados e, portanto, os formulários que se veem de autenticação não utilizam cifra. Portanto, as autenticações vão "em claro" pela rede, permitindo a atacantes, capturar os dados em trânsito.
   - Vamos verificar o que tem a correr: um site que apenas tem autenticação, sem título nem identificação da ferramenta, seu autor, ou outros, típicos para sites web comuns, CMS, etc.

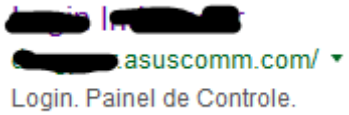
- O facto do “Painel de controle” estar em “pt-br”, confirma a origem do IP brasileiro.
- Foi analisado o código-fonte do site. Os campos também estavam em “pt-br”, tal como “usuário”.

5. O site não tinha título no interface da página de entrada. Não é comum apenas “Painel de controle”. Poderia ser um site feito pelo próprio utilizador. Isso poderia ser útil se quiséssemos introduzir *sql injection* ou explorar vulnerabilidades e tratamento de erros.

Sendo OSINT não pode ser.

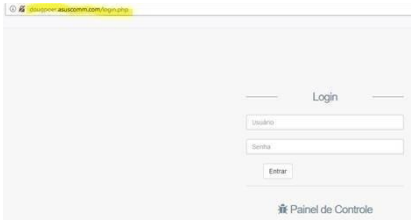
6. Para pesquisar algo único, não podemos ir simplesmente por painel de controlo. É muito comum. Mas no meio do código *html* estava “<title>xxxx xxxx </title>”. Este sim, já é um bom diferenciador e o que nos permitiu avançar.

Foi pesquisado nas redes sociais e nas fontes abertas típicas como motores de busca e agregadores. Procurou-se então por “Painel de Controle” e foram coladas nos motores de busca, o código html parcial dos formulários web.



7. xxxxx? Asuscomm?

Foi feita uma correspondência entre este IP e um nome xxxx.asuscomm.com. Verificou-se que o *dns* acima resolve no mesmo ip e na mesma página anteriormente observada. Não há agora dúvidas.



8. Baseado no nome *asuscomm*, é com muito forte probabilidade que aquele IP seja de um equipamento NAS, mais concretamente um ASUSTOR. (Existe um DNS da ASUS, que pode

ser utilizado pelos NAS da própria ASUS, para que o dono do NAS e quem conhece e ter permissões para aceder ao NAS, não tenha de decorar um IP.)

O atacante tem o site alojado num NAS da ASUS, no sítio de onde lança o ataque? Não sendo muito inteligente, podemos ter aqui duas hipóteses:

a. o atacante ignora que alguém faça o que fiz e faça o portscan sem problemas
b. o atacante seja afinal vitima, e tenha sido o equipamento, parte de *botnet* ou comprometida, a fazer o ataque.

Vamos continuar...

9. Repetindo o M.O., voltou-se a pesquisar e obteve-se o nome *dougpoer*:
    - em vários sites
    - em 1 vídeo no *Youtube* (interessante o vídeo porque o "amigo" xxx xxx ensina a fazer um *portscan*…. Portanto confirma a culpa e o ataque)
    - em sites de jogos. Por exemplo, é jogador de *xxx of xxx*, entre outros
    - em uma universidade, incluindo os seus orientadores de curso
    - usa *linkedin*, onde tem foto, empregos, localizações, etc
    - outros

10. Nesta altura sabemos (*Youtube*) que o *xxx* criou um script de *portscan*, faz publicidade do mesmo no *youtube*, tirou um curso superior de informática, sabemos onde estudou, onde mora, quem foram os seus professores, etc.

No site do *xxxx*, foi possível ver o projecto, assim como no *GitHub*. Vem com isso a identificação, o mail e outros projectos (este é mesmo o único).

Foi descoberto que o atacante:

- chama-se xxx xxx xxx
- programa em *perl*
- tem conta na *sourceforge* onde partilha uma ferramenta que criou
- fez um *crawler* e usa o NAS para recolher e armazenar essa informação

11. Tendo descoberta toda esta informação, foi compilada de forma a traçar um perfil (brincadeira) do "potencial" agressor (última parte deste "*how-to*")

Conclusões do caso prático OSINT

Assim, com “alguma” certeza (nunca total), poderemos ter chegado a partir apenas de um endereço IP, ao seu dono e “potencial” atacante. Tudo foi recolhido com base apenas em OSINT, não ultrapassando qualquer barreira tecnológica nem nunca tendo interagido com o potencial agressor.

Nota: As informações que a seguir são mostradas no “perfil do atacante”, foram obtidas de forma legal, online, sem qualquer autenticação, via IP de VPN e não do IP da instituição onde decorreu a investigação. As informações do perfil do potencial agressor são apenas para leitura e não devem ser usadas de qualquer forma, ou estaríamos também a violar qualquer Lei sobre privacidade, mesmo que tenhamos sido atacados ou o atacante “possa” estar no Brasil.

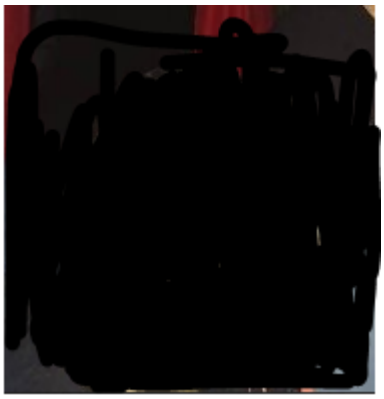
Possivelmente (e ignorando a extensão e a necessidade das forças de segurança em termos processuais e legais), todas as informações que um serviço de informações ou de segurança, poderia precisar para estudar/compreender/deter o *hacker* constam do perfil.

O atacante se fosse minimamente inteligente, teria utilizado uma VPN, não alojaria um site no local de onde lança os ataques, não colocaria o site no seu porto padrão, não colocaria credenciais nem informações pessoais em todo o lado. Tendo em conta a actual profissão, poderemos pensar que foi descuido.

Nota: depois do ataque e recolha de informação nossa no momento, o ip mudou mas o nome http://xxxx.asuscomm.com/login.php continua activo (*portanto teremos sempre o ip mesmo que mude).*

Abaixo segue o perfil compilado.

Relatório - Perfil do atacante



- Identificação: xxx xxxx xxxx
- Origem: xxxxxxxxxxxxx, Brasil
- Emprego: xxxxxx xxxx há 5 anos xxx
- Funções: xxxxx
- Email: xxxxxx.xxxxx@gmail.com
- Capacidades: Programação. xxxxx. Criou um *portscanner* de nome xxxxxx
- Línguas: Segundo o próprio, "xxxx, fala xxxx, escreve xxx lê xxx"
- Estudos:
    - Graduação em xxxxxx xxxx – xxxx
    - Universidade Federal xxxxx, Orientador: xxxx
- Histórico profissional:
    - 2013 - Atual, xxxx
    - 2013 – 2013, xxxxx
    - 2012 – 2013, xxxxx
    - 2011 – 2011, Universidade Federal xxxxx
- Conhecimentos informáticos:
    - Intermédio: C, Delphi, PHP
    - Avançado: Perl
    - Básico: Java

Evidências

- Relatório firewall (que aqui não é disponibilizado) continha o endereço IP e registos ofensivos
- https://www.escavador.com/sobre/xxxx/xxxxx
- https://br.linkedin.com/in/xxxxx
- https://github.com/xxxxx
- https://www.youtube.com/channel/xxxxxx - video
- https://tools.xxxxxx.xxx/web-applications/xxxx - crawler feito
- https://contactout.com/xxxxxxxx
- Foto - https://www.google.com/imgres?imgurl=x%3A%2F%xxxxxxxxxxxxxxxxxxxxxxx.ggpht.com%2Fa-%xxxx-DyUg%3Ds900-mo-c-c0xffffffff-rj-k-no&imgrefurl=xxx_-g_7UOA&docid=ImZL9xqiZYYXtM&xxx=wG5IVvEr4j6aLM%3A&vet=10ahUKEwjXwaqS-7viAhXwUxUIHa__xxx..i&w=900&h=900&itg=1&client=firefox-b-d&bih=1023&biw=2144&q=Douglas%20Poerschke%xxxx&ved=xxxxx-xxx__xxxx&iact=mrc&uact=8

9. Bibliografia e *papers* especializados em OSINT

A bibliografia encontrada relativa a este tema é toda muito recente. O inicio da escrita desta dissertação bate coincidentemente na mesma altura que aumenta a escrita de livros sobre este tema. Embora existam outros, os livros colocados abaixos são dedicados apenas à temática desta dissertação. Todos eles têm data muito recente, o que mostra que o OSINT "está na moda" e que é terreno fértil para investimento em novos estudos e projectos, com muito potencial de retorno em termos de segurança, defesa, combate à cibercriminalidade, conhecimento da concorrência, e auto-conhecimento.

- 2009 - Open Source Intelligence Analysis: A Methodological Approach, Selma Tekir
- 2011 - Counterterrorism and Open Source Intelligence (Lecture Notes in Social Networks), por Uffe Kock Wiil, 29 Jun 2011
- 2013 - 21st Century U.S. Military Documents: Air Force Intelligence and Force Protection (FP), Predictive Battlespace Awareness (PBA), Open Source Intelligence (OSINT), ISR, Contingency Unit Kindle Edition,
- 2014 - Open Source Intelligence in the Twenty First Century (New Security Challenges), Christopher Hobbs
- 2014 - Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information (3rd Edition), por Michael Bazzell
- 2015 - (OSINT) Open Source Intelligence Glossary: Guide to Keywords, Phrases for Improved Internet Research Results, Nicholas Crowder
- 2015 - Automating Open Source Intelligence: Algorithms for OSINT (Computer Science Reviews and Trends), Robert Layton and Paul A Watters, 4 Dec 2015
- 2015 - Automating Open Source Intelligence: Algorithms for OSINT por Dec 2015 por Layton, Robert
- 2015 - *Hacking Web Intelligence: Open Source Intelligence and Web Reconnaissance Concepts and Techniques*, 1 May 2015
- 2015 - Open Source Intelligence Abstraction Layer: Proposta per una Teoria Generale dell'Intelligence delle fonti aperte (Italian Edition) por Giovanni Nacci
- 2016 - (OSINT) Open Source Intelligence Glossary: Guide to Keywords, Phrases for Improved Internet Research Results, Nicholas Crowder
- 2016 - Open Source Intelligence (OSINT) Done Right: An Indictment of 25 years of expensive passive failure., ROBERT David STEELE
- 2016 - Web Intelligence: OSINT Field Guide, De Vlaminck, Bart, 8 Jun 2016
- 2017 - Down the Rabbit Hole an OSINT Journey: Open Source Intelligence Gathering for Penetration Testing, Chris Kubecka, 29 Jun 2017
- 2017 - Open Source Intelligence Investigation: From Strategy to Implementation (Advanced Sciences and Technologies for Security Applications), Babak Akhgar, P. Saskia Bayerl, et al. 9 Jan 2017
- 2018 - Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence, Nihad A. Hassan and Rami Hijazi
- 2018 - Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information, Nihad A. Hassan and Rami Hijazi – 26 Jan 2018

- 2019 - Hack The World with OSINT (Hackers Gonna Hack), Chris Kubecka and Martinez II, E , 17 Jan 2019
- 2019 - OSINT for the Staffing World! Da Costa, Dean
- 2020 - Hunting Cyber Criminals: A Hacker's Guide to Online Intelligence Gathering Tools and Techniques, 3 Feb 2020

Bibliografia

Appel, Edward J.; Internet Searches for Vetting, Investigations, and Open-Source Intelligence, disponível online em https://books.google.pt/books?id=hIq6ku99UDYC&pg=PA178&lpg=PA178&dq=google+como+fazer+open+source+intelligence&source=bl&ots=ZT63BgYG7B&sig=p03RyzOLB9F-2FU9zjymBjQgIVw&hl=pt-PT&sa=X&ei=4FY-VdWlKcbl7gbx6oDwBg&ved=0CDIQ6AEwAg#v=onepage&q=google%20como%20fazer%20open%20source%20intelligence&f=false

Arthur S. Hulnick, Capítulo "The Dilemma of Open Sources intelligence: Is OSINT Really Intelligence? ", The Oxford Handbook of National Security Intelligence, editora Loch K. Johnson

Bazzell, Michael; Open Source Intelligence Techniques, 3rd Edition, 2014

Best, Richard A., Jr e Cumming, Alfred; Open Source Intelligence (OSINT): Issues for Congress, 2007.

Bravo, Rogério - "Open Sources" na investigação do cibercrime: conceito e implicações, 2014, acedido em 2015/12/12 em https://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime

Bravo, Rogério e Góis, Nuno; Por um um novo modelo de análise integrada em estudos de intelligence e de estratégia, Lisboa, 27 de Junho de 2011

Bravo, Rogério, Por um novo modelo de análise holística em estudos de intelligence, v. 3 de 19 de Julho de 2013

Bravo, Rogério; O Conceito de "Fontes Abertas " na Investigação do Cibercrime https://www.academia.edu/5906230/O_Conceito_de_Fontes_Abertas_na_Investiga%C3%A7%C3%A3o_do_Cibercrime, 2014

Bravo, Rogério; "OPEN SOURCES IN CYBERCRIME INVESTIGATION: concept and implications"

Bravo, Rogério; Iniciativas Legais de Combate ao Cibercrime e às Ciberameaças

Bravo, Rogério; OPEN SOURCES IN CYBERCRIME INVESTIGATION: concept and implications, Policia Judiciária, disponível em http://www.academia.edu/7301278/OPEN_SOURCES_IN_CYBERCRIME_INVESTIGATION_concept_and_implications

Bravo, Rogério; Por um novo modelo de análise holística em estudos de intelligence, v. 3 de 19 de Julho de 2013

Bravo, Rogério; Por um novo modelo de análise integrada - estudos de intelligence para os sete espaços, lisboa, 27 de junho de 2011

BRELSFORD, PAUL. Employing a social media monitoring tool as an OSINT platform for Intelligence, Defence & Security, acedido a 19-02-2015 e disponível online em https://www.eurosint.eu/system/files/employing_social_media_monitoring_tools_as_an_osint_platform_for_intelligence_defence_security.pdf

brito, p. (2015, 02 18). Google indexa 650 mil páginas/dia na Dark Web. Obtido a 02 19, 2015, de Cybersecurity.com: http://www.cibersecurity.com.br/google-indexa-650-mil-paginasdia-na-dark-web/

Casanovas, Pompeu; Open Source Intelligence, Open Social Intelligence and Privacy by Design, http://ceur-ws.org/Vol-1283/paper_24.pdf

Delong, Maxence; Filiol, Eric; Coddet, Clément; Fatou, Olivier; Suhard, Clément, ESIEA Laval, Paper, OSINT Analysis of the TOR Foundation, disponível online em https://arxiv.org/pdf/1803.05201.pdf

Electronic Intelligence at NSA, National Security Agency, visto a 2016/02/09 e disponível online em https://www.nsa.gov/about/_files/cryptologic_heritage/publications/misc/elint.pdf

equation group: questions and answers, http://paper.li/dfern__/1392983556

FREITAS, Daniel Ferreira COSTA, Vinicius Rodrigues da; O USO DE FONTES ABERTAS NO TRABALHO DE INTELIGÊNCIA POLICIAL, paper, Academia da Polícia Militar de Goiás, disponível online em

Grey, Steph, Agentes secretos, a nova espionage mundial, Editora Clube do Leitor

https://acervodigital.ssp.go.gov.br/pmgo/bitstream/123456789/1204/1/Daniel%20Ferreira%20Freitas.pdf

https://comum.rcaap.pt/bitstream/10400.26/8749/1/o%20papel%20dos%20servi%C3%A7os%20de%20informa%C3%A7%C3%B5es%20no%20combate%20ao%20ciberterrorismo%20O%20caso%20Portugu%C3%AAs%20..pdf

Hulnick, Arthur S. The Dilemma of Open Sources intelligence: Is OSINT Really Intelligence? The Oxford Handbook of National Security Intelligence

Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT) por Michael Glassmana, e Min Ju Kang, disponível online em http://www.sciencedirect.com/science/article/pii/S0747563211002585

Johnson, Loch K.; Handbook of Intelligence Studies

Koops, Bert-Jaap, Police Investigations in Internet Open Sources: Procedural-Law Issues, Tilburg University, TILT. Paper disponível online em https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2574951

Manual de Informações, Lisboa, Estado-Maior do Exército, 1979, referenciado na página 54

Mercado, Stephen C.; Sailing the Sea of OSINT in the Information Age

Military Intelligence, Tenente General Vizela Cardoso, disponível online em http://www.afceaportugal.pt/docs/viii_symposium/Microsoft_PowerPointVCARDOSO_-_1.14_.pdf. Acedido a 29-04-2015

O Papel dos Serviços de Informações no Combate ao Ciberterrorismo, Sandra Núria Basto Perez do Amaral, tese de mestrado em Guerra da Informação, Academia Militar, 2014

Olcott, Anthon; Open Source Intelligence in a Networked World, https://books.google.pt/books?id=dHTo5g7QYEUC&pg=PT6&lpg=PT6&dq=google+como+fazer+open+source+intelligence&source=bl&ots=6R4xLnqM44&sig=1aGyOsTrEFlU-jb7HfwH_3QJ25g&hl=pt-PT&sa=X&ei=4FY-VdWlKcbl7gbx6oDwBg&ved=0CDcQ6AEwAw#v=onepage&q=google%20como%20fazer%20open%20source%20intelligence&f=false

Open Source Intelligence (OSINT): Issues for Congress, Richard A. Best, Jr e Alfred Cumming. 5/12/2007. Disponível online em http://www.fas.org/sgp/crs/intel/RL34270.pdf

Relatório Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution, acedido a 4-1-2018 em https://www.dni.gov/files/documents/ICA_2017_01.pdf, National Intelligence Council EUA

Russel, Mathew A. X; Oxford: Oxford University Press.Mining the social web, O'Reilly, 2011

Schaurer, Florian e Störge, Jan; The Evolution of Open Source Intelligence (OSINT), Winter/Spring 2013

Schaurer, Florian e Störger, Jan; The Evolution of Open Source Intelligence (OSINT), disponível online em http://www.afio.com/publications/Schauer_Storger_Evo_of_OSINT_WINTERSPRING2013.pdf

SERRA, Pedro Paulo José de Sousa Teles, Sistema de Informações Militares. Contributos para a sua reestruturação e operacionalidade. Disponível online em https://comum.rcaap.pt/bitstream/10400.26/11894/1/MAJ%20Serra%20Pedro.pdf, referenciado na página 61

Steele, R. (n.d.). Forbes. Retrieved from http://www.forbes.com/2006/04/15/open-source-intelligence_cx_rs_06slate_0418steele.html

Stephen C. Mercado. Sailing the Sea of OSINT in the Information Age, disponível online em https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol48no3/pdf/v48i3a05p.pdf

The Evolution of Open Source Intelligence (OSINT), Florian Schaurer and Jan Störge, Winter/Spring 2013, disponível online em http://www.afio.com/publications/Schauer_Storger_Evo_of_OSINT_WINTERSPRING2013.pdf

Wirtz, J. (2010). The sources and methods of intelligence studies. In JohnsonL. (Ed.), The Oxford Handbook of National Security Intelligence(pp.59-69).